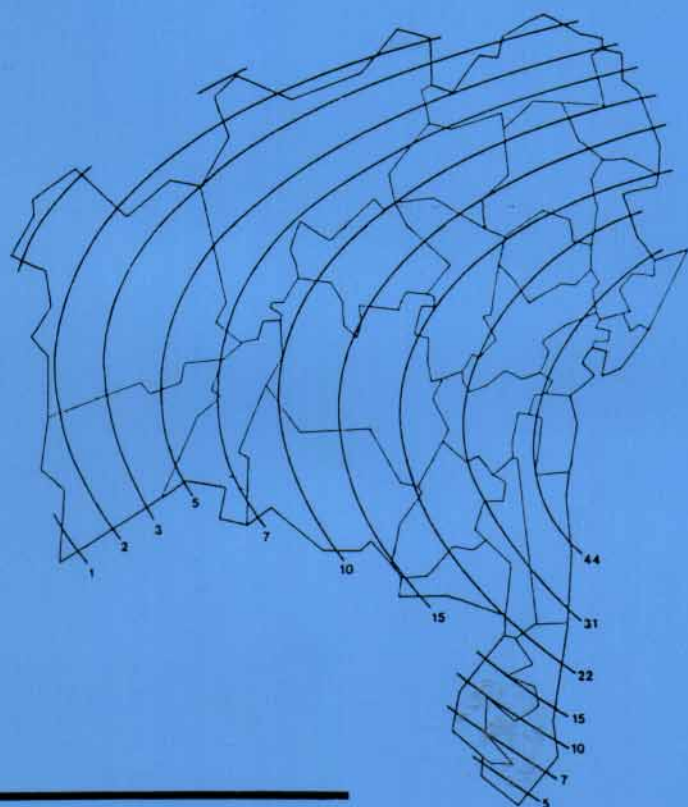


# LES DONNÉES ET LE TERRITOIRE

initiation  
à l'analyse  
en surfaces  
de tendance



Philippe Waniez  
Yann Le Gauffey

Editions de l'ORSTOM

INSTITUT FRANCAIS DE RECHERCHE SCIENTIFIQUE POUR LE DÉVELOPPEMENT EN COOPÉRATION

# **LES DONNÉES ET LE TERRITOIRE**

## **initiation à l'analyse en surfaces de tendance**

**Philippe Waniez**

Géographe ORSTOM

**Yann Le Gauffey**

*avec la collaboration de*

**Violette CABOS**

Ingénieur cartographe au CNRS

---

**Éditions de l'ORSTOM**

INSTITUT FRANÇAIS DE RECHERCHE SCIENTIFIQUE POUR LE DÉVELOPPEMENT EN COOPÉRATION

Collection **INITIATIONS - DOCUMENTATIONS TECHNIQUES** n° 69

PARIS 1987

Ce fascicule a été rédigé par Philippe Waniez, géographe chargé de recherche à l'ORSTOM à Brasilia, et Yann Le Gauffey, docteur en géographie,

avec la collaboration de Violette Cabos, qui a assuré la cartographie et la composition du texte à la Maison de la Géographie de Montpellier (G.I.P. RECLUS),

dans le cadre de la Convention ORSTOM-EMBRAPA (projet SISECSO)

La loi du 11 Mars 1957 n'autorisant, aux termes des alinéas 2 et 3 de l'article 41, d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective » et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation ou reproduction intégrale, ou partielle, faite sans le consentement de l'auteur ou de ses ayants droit ou ayant cause, est illicite » (alinéa 1<sup>er</sup> de l'article 40).

Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles 425 et suivants du Code Pénal.

## 0. SURF: un progiciel d'analyse en surfaces de tendances.

Les ouvrages de statistique usuelle n'abordent que très rarement le problème de la recherche des tendances spatiales; on y trouve une grande quantité de techniques, souvent complexes, de mesures de tendances sur des données temporelles, du type série économique par exemple. Pour les données spatialisées, il faut aller à la recherche d'une documentation plus spécialisée dont font état les bibliographies d'ouvrages d'introduction aux méthodes quantitatives en géographie, qui ne font, la plupart du temps, qu'effleurer la question. S'il ne possède pas une solide formation en mathématiques, le géographe renoncera assez vite, découragé par une formalisation à laquelle il est peu habitué. Même s'il arrive à assimiler les aspects techniques de ces méthodes, il lui restera à concevoir et programmer l'outil informatique indispensable au traitement des données.

Heureusement, plusieurs auteurs, géographes et renommés pour la qualité de leurs travaux méthodologiques, ont compris que l'excellence d'une méthode n'en induisait pas forcément sa bonne diffusion, ni une large expérimentation nécessaire à son utilisation optimale (en fonction des limitations apparues). Parmi ces auteurs, trois retiennent l'attention pour l'intérêt et la clarté de leurs exposés méthodologiques, mais aussi et surtout pour avoir publié leurs programmes informatiques. Ainsi, le présent fascicule reprend une part importante des travaux de:

LEMAY (G.) - 1984 -  
L'étude de la tendance spatiale.  
Université Paris X-Nanterre  
Informatiques-Informations

MATHER (P.M.) - 1976 -  
Computational Methods of Multivariate  
Analysis in Physical Geography.  
Chichester, Wiley, 531 p.

WRIGLEY (N.) - 1977 -  
Probability Surface Mapping  
Norwich, CATMOG n°16, 78 p.

Après avoir tenté de faire fonctionner ces programmes publiés sous forme imprimée (ce qui n'est pas toujours aisé en raison de la multiplicité des dialectes FORTRAN), et perçu l'immensité du champ d'application, un premier effort d'information s'est conclu par la publication par l'Université Paris X-Nanterre d'un numéro spécial de la revue Informatiques-Informations, intitulé "SURFACES DE TENDANCES" (n° 5, 1984).

Quelques temps plus tard, la diffusion des micro-ordinateurs IBM PC et compatibles dans les universités et les centres de recherche incite à poursuivre et étendre ce dernier effort en fournissant le système qui faisait défaut jusqu'à présent.

Le progiciel SURF est un système informatique d'analyse en surfaces de tendances. Il propose les trois techniques programmées par Lemay, Mather et Wrigley. Les programmes ont été adaptés au FORTRAN 77 Microsoft et sont fournis sur disquette sous forme de modules immédiatement exécutables sur tout micro-ordinateur IBM PC ou 100 % compatible.

L'ouvrage est organisé de la manière suivante: le premier chapitre présente les bases méthodologiques de l'analyse en surfaces de tendances et discute le concept de composante d'échelle; cet exposé s'appuie sur deux exemples brésiliens. Le chapitre suivant explique ce qu'est une matrice d'information tendancielle et comment il faut la construire. Le chapitre 3 est réellement le cœur du livre; on y présente successivement et avec un certain détail les trois techniques retenues ici. Sont également exposés l'utilisation du progiciel et le commentaire des sorties d'ordinateur. Enfin, une annexe traite de l'installation du produit sur micro-ordinateur IBM PC ou compatible à l'aide des disquettes fournies en option avec le livre.

Le but de cet ouvrage n'est pas de couvrir la totalité de la question de la recherche des tendances en géographie. En particulier, le problème des tendances périodiques

(notamment celui de l'analyse spectrale spatiale) n'est pas traité; il serait intéressant de l'inclure au progiciel SURF et les auteurs d'un tel programme sont cordialement invités à nous contacter. La présente introduction pratique se veut utile à tous ceux qui souhaitent progresser en expérimentant une méthode encore trop peu diffusée.

Les auteurs:

Philippe WANIEZ

Docteur en géographie.

Chargé de recherche à l'ORSTOM.

Responsable français d'un programme de recherche franco-brésilien pour la réalisation d'une base de données économiques et sociales relative au développement de l'agriculture dans la région des Cerrados (programme SISECSO).

Yann LE GAUFFEY

Docteur en géographie.

A collaboré durant de nombreuses années au développement de l'informatique appliquée aux sciences sociales à l'Université Paris X-Nanterre. Chef de projet informatique dans une grande entreprise industrielle.

---

## 1. L'analyse en composantes d'échelles et la cartographie des tendances

Véritable méthode d'analyse des structures de l'espace géographique, l'analyse en surfaces de tendances (AST) a un domaine d'application immense, recoupant toutes les branches de la géographie; qu'elle soit physique ou régionale, la recherche géographique anglo-saxonne a abondamment fait appel, depuis plus d'un quart de siècle, à cet ensemble de techniques également connu sous le nom de cartographie des surfaces de tendances, la carte constituant à la fois le matériau et le produit de l'analyse. Moins pratiquée par les chercheurs francophones, probablement occultée par l'Analyse des Données et en particulier l'analyse factorielle, l'AST est à proprement parler une méthode d'analyse en composantes d'échelles où localisations et distances jouent un rôle prépondérant. Par composantes d'échelles, il faut entendre l'existence dans un phénomène de dimensions résultant chacune de jeux d'influences qui s'exercent sur des étendues de plus en plus petites. Ainsi, en présence de l'organisation d'un espace donné, on peut faire l'hypothèse du déroulement de processus hiérarchisés, 1/ processus généraux affectant non seulement la totalité de l'espace étudié, mais également ses voisins, même éloignés, 2/ processus régionaux couvrant toute l'aire d'étude, 3/ processus locaux ne concernant qu'une portion seulement du territoire. A chacune de ces échelles correspond une composante qui, une fois cartographiée, prend la forme d'une surface de tendance. La parenté avec l'analyse en composantes principales est évidente: on cherche, au travers du foisonnement des données, les quelques dimensions qui rendent compte, au mieux, des variations observées, tout en estimant leur degré de généralité; ici, cependant, ces dimensions n'ont pas l'apparence d'un axe, mais d'une surface. Bien que proche d'une simple technique de lissage sur le plan des calculs mis en oeuvre, l'AST s'en éloigne en raison de son caractère itératif: on ne cherche pas directement le meilleur ajustement, mais on examine la signification des étapes successives qui permettent de l'atteindre.

Présenter quelques exemples constitue un excellent moyen de compréhension d'une méthode. Parmi toute une série d'expérimentations réalisées à cette fin, deux ont été retenues, intéressantes sur le plan géographique, et pouvant de plus faire l'objet d'un exposé didactique. Le lecteur enclin à formuler des critiques sur le fond, c'est-à-dire sur les résultats géographiques à proprement parler, ne doit pas s'attendre à trouver ici de grandes découvertes scientifiques; il s'agit d'un texte qu'on a voulu aussi simple que possible, les aspects plus complexes et plus techniques venant dans les chapitres ultérieurs.

### 1.1. Un exemple élémentaire: les densités de population dans l'état de Bahia au Brésil

Situé dans la région Nordeste du Brésil, l'état de Bahia est caractérisé, sur le plan des densités démographiques, par un très grand contraste entre le littoral et l'intérieur. En effet, c'est sur la côte atlantique que sont localisés les deux principaux foyers de peuplement (fig. 1.1). D'une part, la capitale, Salvador, forme une grande agglomération de deux millions d'habitants et s'étend, autour de la Baie de Tous les Saints, sur un rayon de plus de 100 kilomètres dans le *Reconcavo Baiano*; les densités y sont toujours supérieures à 80 habitants par kilomètre carré et dépassent 4 000 dans le *município* de Salvador. D'autre part, plus au sud, la grande région cacaônière centrée sur les villes d'Itabuma et Ilheus (respectivement 150 000 et 80 000 habitants) a une densité moyenne approchant les 90 habitants au kilomètre carré. Au delà de ces deux foyers de peuplement, les densités diminuent très rapidement: c'est le domaine de la *caatinga* au climat semi-aride caractéristique du *Sertão Nordestin*. On y rencontre le plus souvent des valeurs comprises entre 25 et 50 habitants au kilomètre carré. Encore plus à l'ouest, les *cerrados* n'ont, sauf quelques exceptions, qu'une densité le plus souvent inférieure à 5 habitants au kilomètre carré. Ainsi, la densité de population dans l'état de Bahia présente une configuration spatiale simple: de fortes valeurs sur la côte, diminuant vers l'ouest pour atteindre quelques habitants au kilomètre carré seulement à l'extrême ouest.

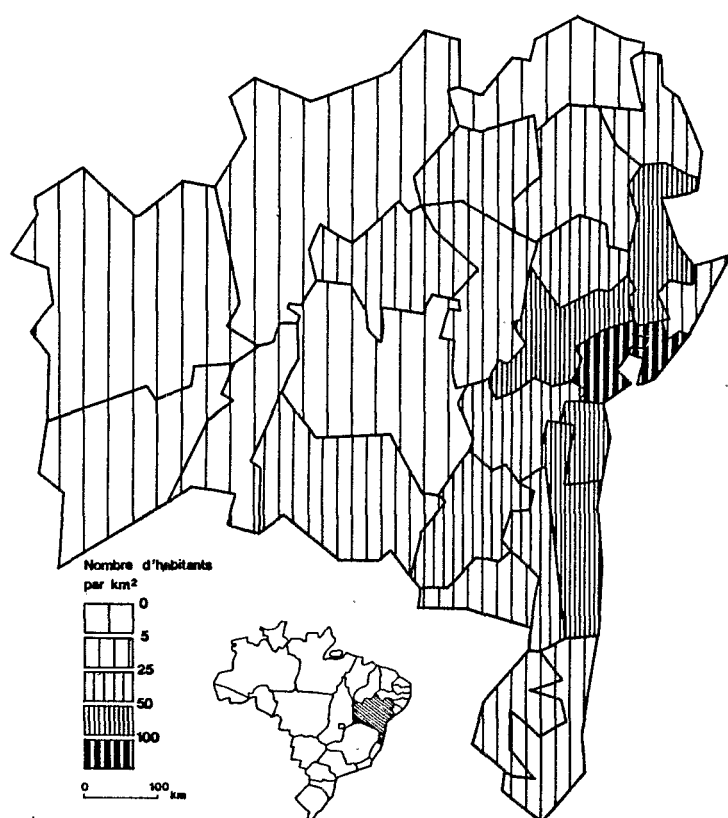


fig. 1.1 La densité de population dans l'état de Bahia (micro-régions, 1980)

L'AST des données statistiques ayant permis de dresser la précédente carte des densités de population conduit, bien entendu, à des observations semblables. En effet, s'agissant des mêmes données mesurant le même phénomène dans le même espace, il n'y a aucune raison d'obtenir des résultats profondément différents, surtout dans le cas présent, si simplement et fortement structuré dans l'espace. Une première surface (fig. 1.2) représente la décroissance des densités de l'est vers l'ouest sous forme de lignes d'iso-densité parallèles et légèrement inclinées dans la même direction que la côte Atlantique. La figure 1.3 nous apprend qu'en règle générale, au Brésil, la densité de population diminue en fonction de l'éloignement de la côte. Bahia ne fait pas exception à cette règle. La composante d'échelle figurée par la première surface traduit un phénomène d'ampleur nationale. Un indicateur statistique élémentaire, le taux de variation expliquée, confirme cette observation par sa valeur significative. La seconde surface (fig. 1.4) a une forme plus complexe: les lignes d'iso-densité s'organisent selon une demi-ellipse, dont l'un des foyers est localisé à proximité de la capitale, Salvador. Cette surface cumule le phénomène général de décroissance des densités de la côte vers l'intérieur, avec une tendance particulière à Bahia, tendance que l'on peut qualifier de régionale pour simplifier la présentation. Le

taux de variation expliquée confirme ce caractère plus régional. Ainsi, la position de Salvador, au sommet de la surface des densités, induit une interprétation en terme de polarisation; en plus de la distance à la côte, l'éloignement de la capitale renforce les disparités de l'occupation humaine.

Deux composantes d'échelle permettent de produire une carte reflétant assez bien la carte d'origine. Ce modèle, obtenu par le calcul, a le mérite de démontrer, sans recourir à une généralisation intuitive, l'existence de deux niveaux d'analyse, national et régional, et d'en apprécier le poids respectif. Cependant, cette étude n'est réellement achevée qu'avec l'examen des déviations par rapport à l'idéal figuré par le modèle, faciles à estimer à partir des équations représentatives des surfaces. La figure 1.5 est une carte de ces déviations. On y découvre la présence de plusieurs cas d'inadéquation, localisés en majorité dans la partie est de la carte, où se situent précisément les plus fortes densités. On distingue des déviations positives, là où le modèle sous-estime les valeurs par rapport à la réalité, et négatives dans les cas de surestimation.

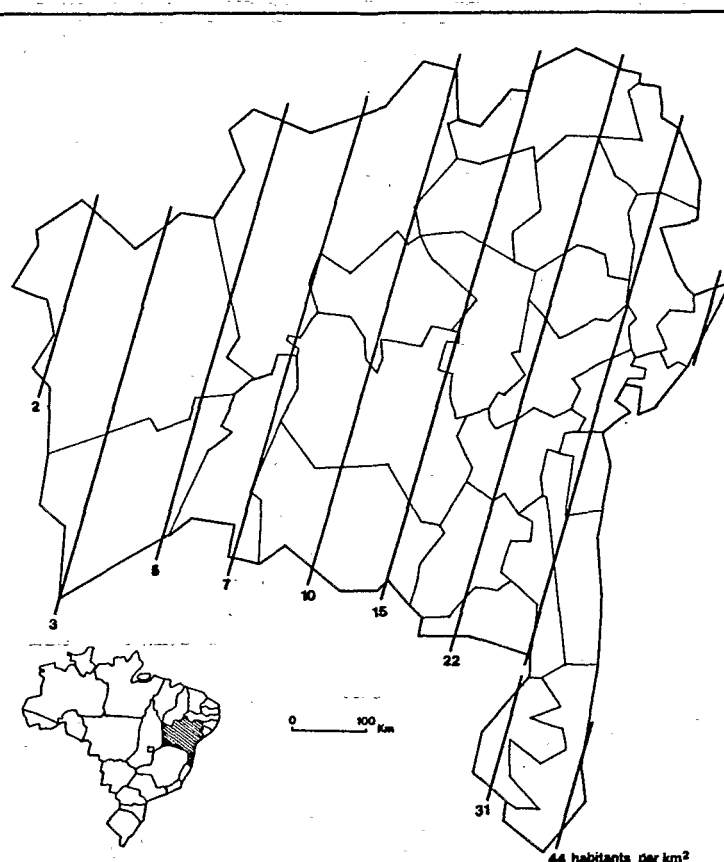
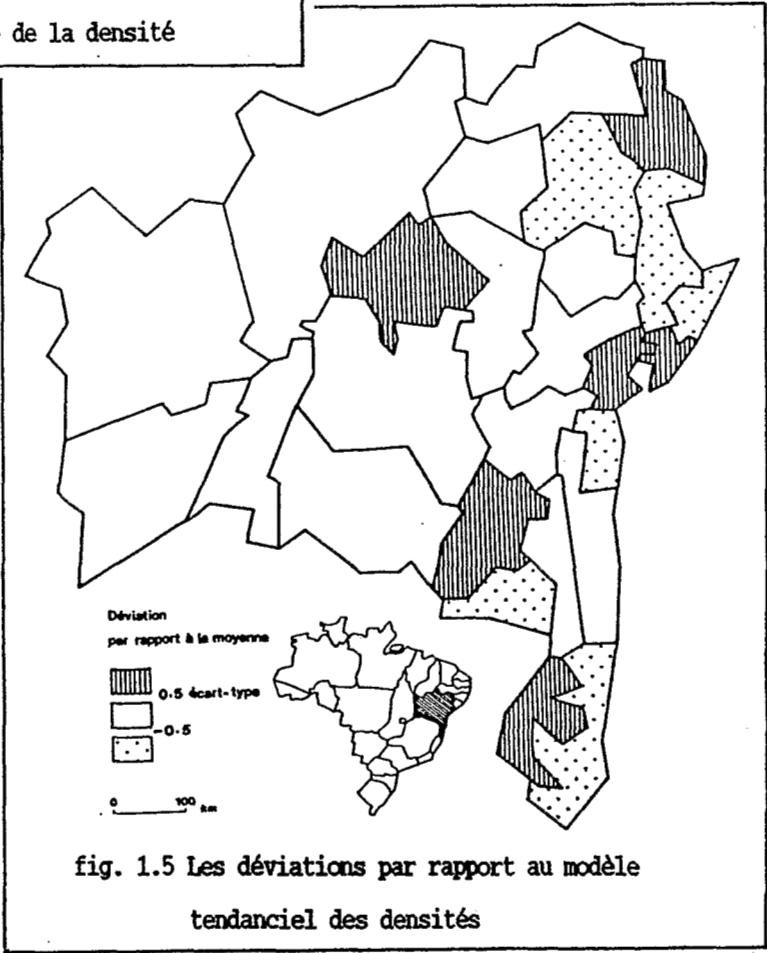
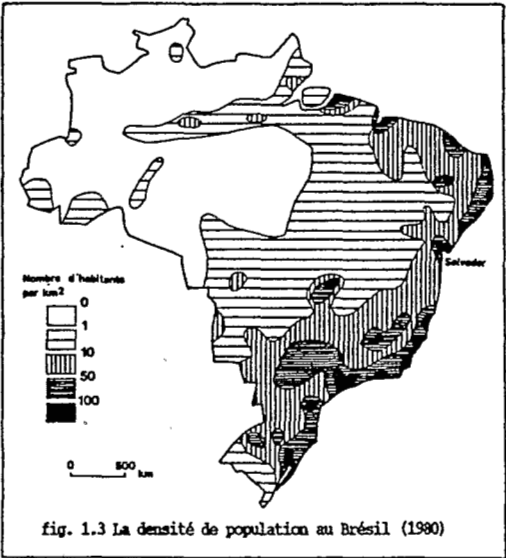
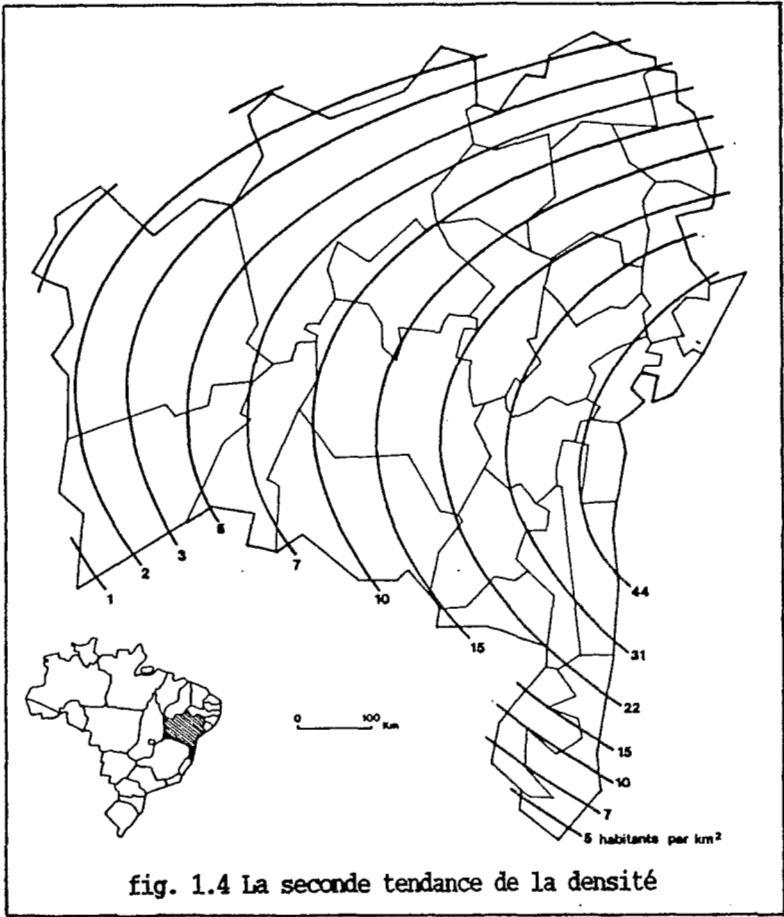


fig. 1.2 La première tendance de la densité





Partout, il s'agit de situations particulières qu'il est vain de chercher à expliquer par une tendance, mais dont la mise en évidence constitue un fait géographique en soi. Salvador et le Reconcavo Baiano, par exemple, tous deux sous-estimés, sont une donnée essentielle de la géographie de l'état de Bahia.

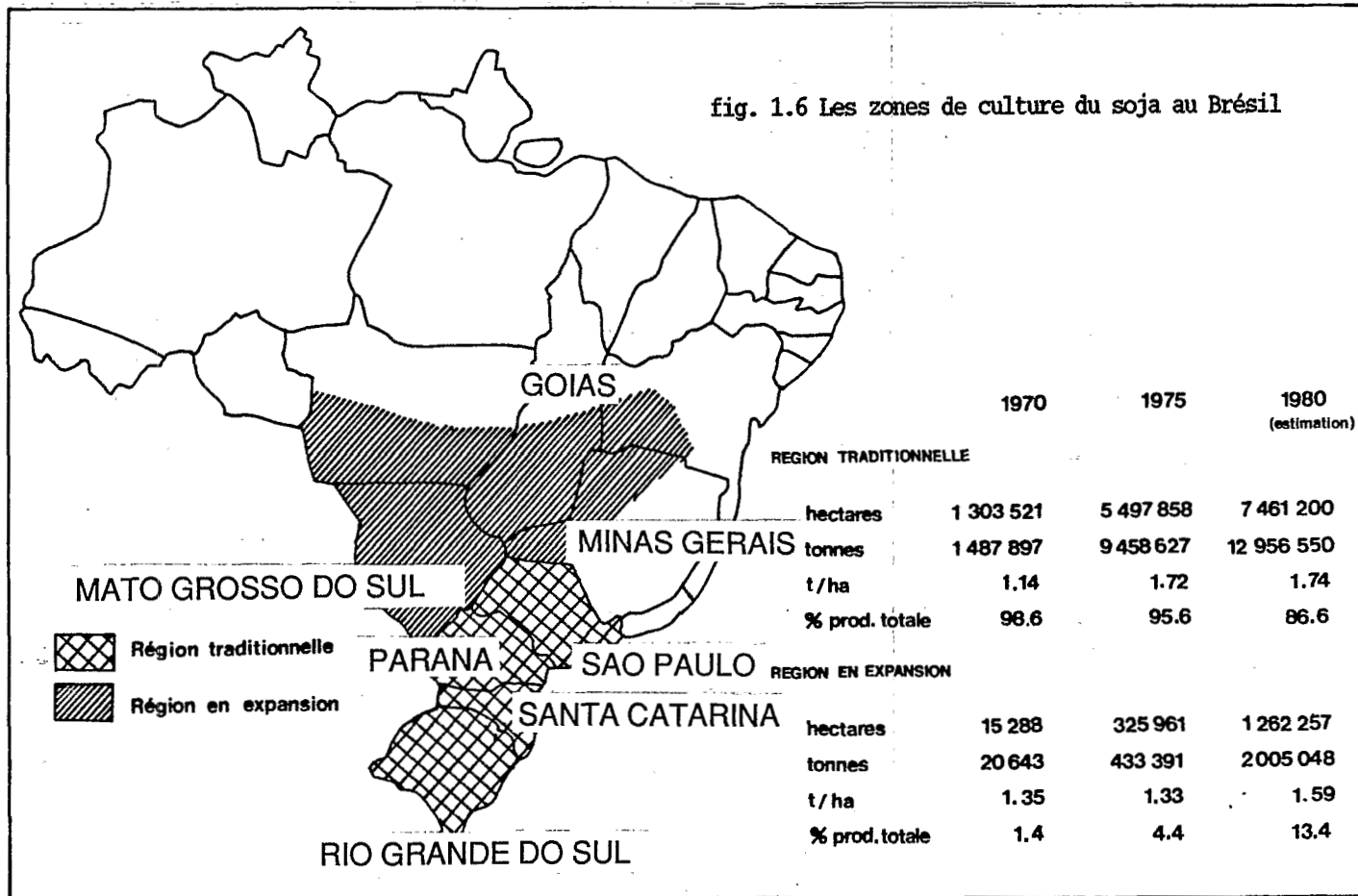
## 1.2. Un exemple plus complexe: l'expansion de la culture du soja au Brésil

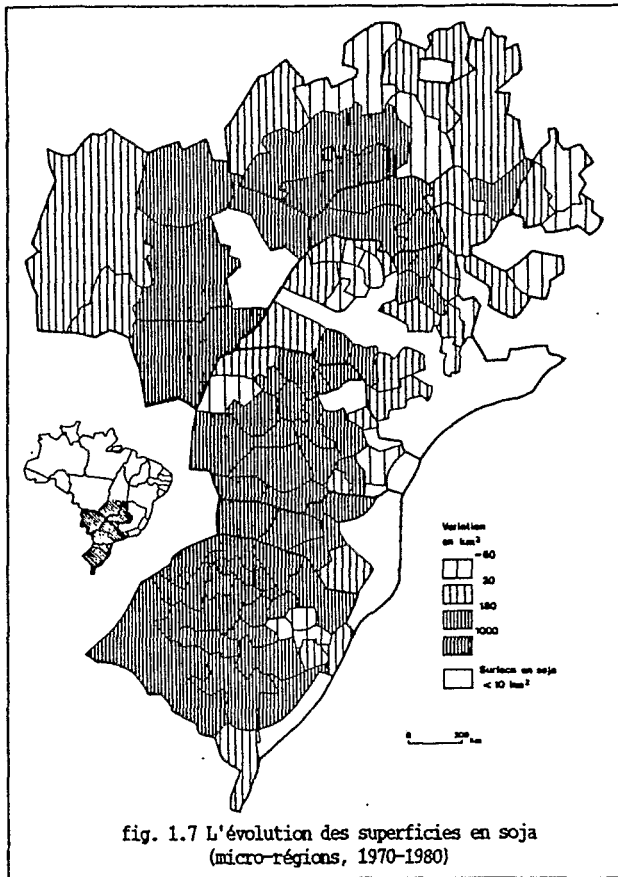
Introduit au Brésil, à titre expérimental, au début du XIX<sup>e</sup> siècle, le soja n'y a atteint un petit volume de production significatif que vers les années 1950. L'essentiel était localisé dans le Rio Grande do Sul, avec quelques tentatives dans les états de Sao Paulo et du Parana. C'est surtout pendant la période 1970-1980 qu'on a enregistré une production importante sur le plan économique: elle est passée de 1.5 à 15 millions de tonnes avec un taux de croissance annuel de 30 %; ainsi, le soja représentait en 1980 1/5 de la superficie cultivée du Brésil, avec 8.7 millions d'hectares. Les principaux facteurs ayant contribué à cette expansion sont les suivants: 1/ conditions favorables du marché, 2/ Possibilité d'assolement avec le blé permettant de faire deux récoltes par an, 3/ facilité de mécanisation, 4/ capitalisation des entreprises agricoles.

En fonction de l'époque du développement initial de la culture, on distingue deux zones géographiques. La zone traditionnelle couvre les états du Rio Grande do Sul, Santa Catarina, Parana et Sao Paulo (fig. 1.6). En 1980, on y trouvait près 87 % de la superficie en soja, avec un niveau technologique élevé et un rendement moyen de 1740 kg/hectare. La zone d'expansion couvre le Mato Grosso do Sul, la partie sud du Goias et du Mato Grosso, l'ouest du Minas Gerais et de la Bahia. Ici, la principale phase de développement se situe entre 1975 et 1980, la superficie ayant quadruplé en cinq ans. En raison notamment de conditions édapho-climatiques plus difficiles, et malgré un niveau technologique élevé, la production moyenne en 1980 est de 150 kg/hectares inférieure à celle du sud. Un gros effort est fait pour améliorer ces performances en produisant des génotypes plus adaptés et en mettant en place des systèmes de culture plus efficaces.

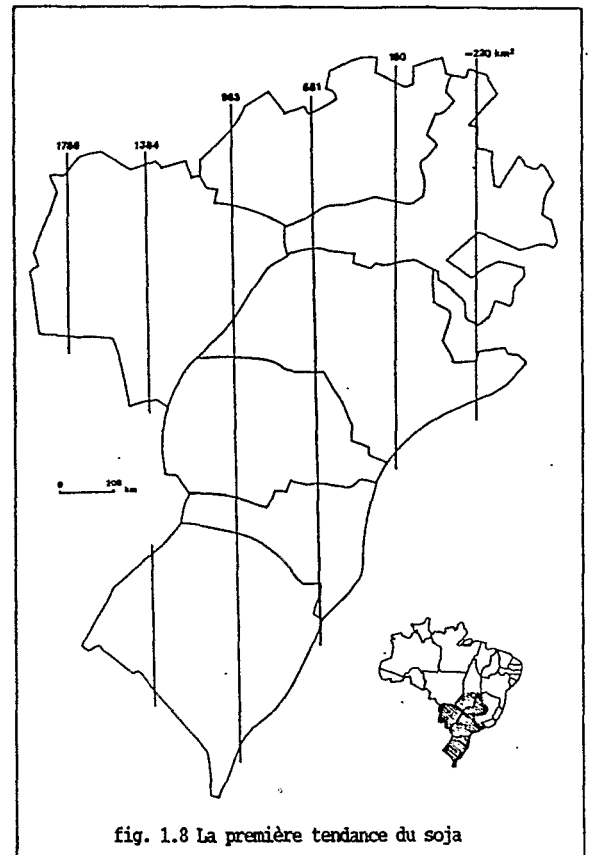
La carte, établie au niveau micro-régional, de l'évolution des superficies en soja précise ce qui vient d'être observé sur les états (fig. 1.7). On voit très bien se dessiner les principales zones de développement de cette culture: les bordures ouest de la zone traditionnelle (du Rio Grande do Sul jusqu'au Parana), la partie centrale du Mato Grosso do Sul et le sud du Goias, dans la zone d'expansion.

fig. 1.6 Les zones de culture du soja au Brésil



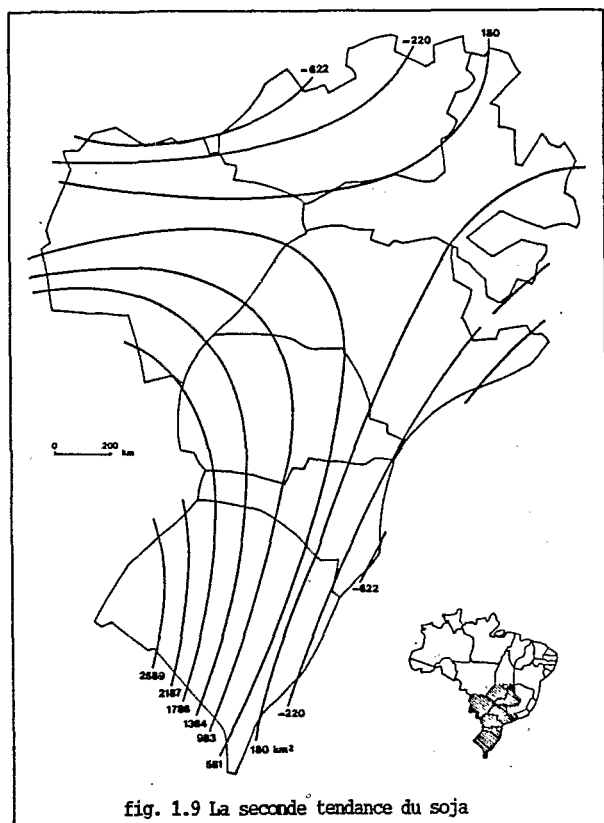


Les résultats de l'analyse en surfaces de tendances de la matrice de données (SOJA7080, voir chapitre 2) à la base de la précédente carte, sont riches d'enseignements. La première surface est figurée par une série de lignes parallèles, d'orientation nord-sud, représentant l'accroissement des superficies en soja de l'est vers l'ouest (fig. 1.8). La vocation de cette plante à être une culture de l'intérieur se trouve donc confirmée ici. On doit alors se demander en quoi cette orientation méridienne représente une composante d'échelle générale, dépassant le cadre de l'aire d'étude, ce qui, rappelons-le, est la principale caractéristique de la première surface. Sachant que le soja vaut la peine d'être cultivé dans le cadre de grandes exploitations agricoles, disposant d'une mécanisation très poussée, il est clair qu'il se développe d'autant mieux qu'il existe de grands espaces vacants aptes à la mécanisation. Ceci étant l'une des caractéristiques de l'intérieur du Brésil (le Planalto Méridional est la principale zone d'extension du soja), mise à part la forêt amazonienne, il est normal que la première composante d'échelle exprime la combinaison de ces facteurs, conditionnant la progression géographique du modèle de développement agricole symbolisé par le soja. Le taux de variation expliquée nous enseigne que la tendance qui vient d'être mise à jour est très significative, et n'est donc pas seulement le résultat d'un procédé de calcul.



La seconde tendance, également très significative, apporte une information supplémentaire intéressante (fig. 1.9). Elle met en évidence une limitation d'orientation zonale: la progression du soja vers le nord est de plus en plus difficile, surtout au-delà du Parana. On voit apparaître d'une part, la zone de culture traditionnelle, dans la partie méridionale, avec une expansion d'orientation méridienne, et d'autre part, la zone de diffusion vers le nord. Au sens large du terme, c'est bien d'un phénomène "régional" dont il s'agit, plus particulièrement lié aux contrastes édapho-climatiques entre les parties sud et nord de l'aire d'extension de la culture, la seconde présentant plus de difficultés culturelles. Remarquons enfin la présence d'un "col" compris entre 580 et 180 kilomètres carrés, centré sur le Triangle Mineiro (Uberlândia, Uberaba), l'un des principaux pôles d'expansion.

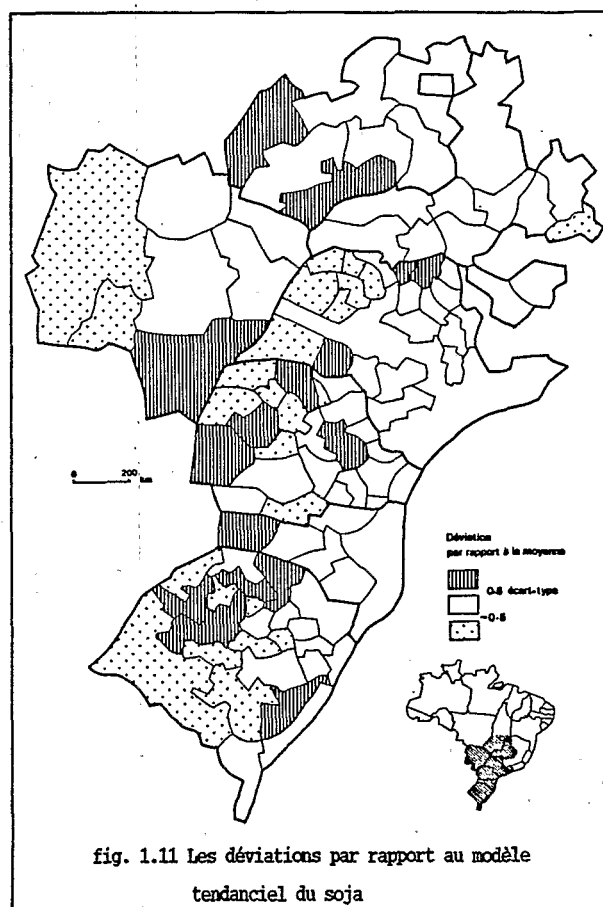
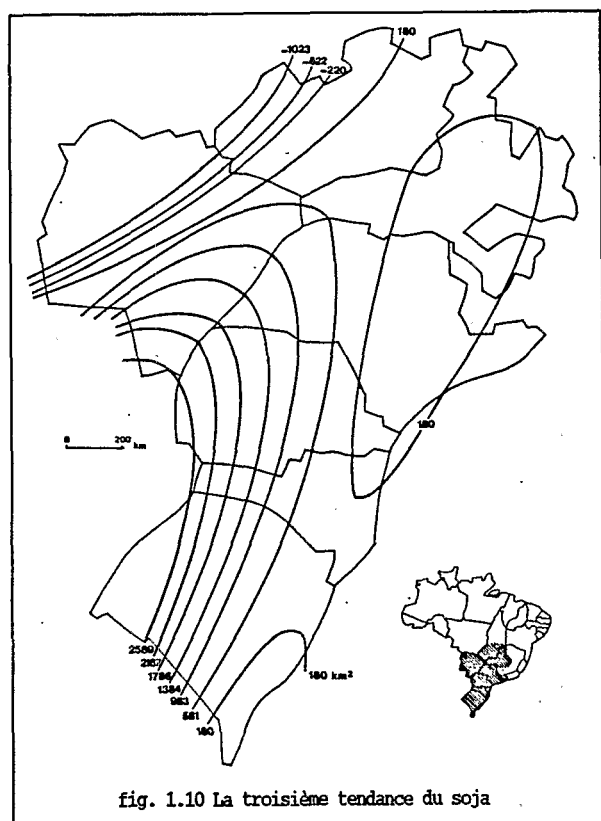
La troisième surface n'apporte qu'un gain d'explication assez réduit; la tendance précédente se trouve de ce fait peu modifiée (fig. 1.10). Le seul élément intéressant, de nature locale, est la présence d'un minimum inférieur à 180 kilomètres carrés, centré sur la frontière séparant les états de Sao Paulo et du Minas Gerais. Cette "dépression" est donc moins favorable à la diffusion du soja que le sud-ouest du Brésil, mais aussi plus favorable que le nord-ouest. Sur ce terrain s'exprime une



concurrence avec d'autres cultures (le café en particulier), et d'autres structures d'exploitations agricoles.

En résumé, l'expansion agricole du soja est assez bien représentée par trois composantes d'échelle, chacune déterminée par des facteurs géographiques de nature différente et s'étendant sur des espaces de plus en plus réduits. La carte des déviations par rapport à ce modèle explicatif (fig. 1.11) ne présente aucun phénomène de coalescence ayant une certaine ampleur géographique; ceci confirme la validité de l'analyse: la méthode de recherche et les résultats qui en découlent présentent donc un intérêt certain.

Ces exemples choisis pour leur caractère démonstratif immédiat, ne doivent pas faire penser que la mise en pratique de l'AST soit toujours aussi simple. Par exemple, dans le cas d'un phénomène de polarisation, la tendance générale ne sera probablement pas représentée par la première surface, mais sûrement par la seconde. Le taux de variation expliquée doit faire l'objet, dans tous les cas, d'un examen attentif.



## 2. Elaborer une matrice d'information tendancielle

Préalablement à toute analyse, il est indispensable d'organiser les données sous la forme d'un tableau. Dans le cas de l'analyse en surfaces de tendances, il s'agit d'une variété particulière du genre matrice d'information spatiale, nommée matrice d'information tendancielle, résultant de la mise en correspondance de deux ensembles: celui des unités spatiales et celui des attributs géographiques

### 2.1. Les unités spatiales

Elles forment un semis de points dans l'espace, même lorsqu'il s'agit des éléments d'une maille administrative; dans ce dernier cas, il faut choisir le point le plus représentatif, ce qui ne va pas toujours sans difficulté. Chaque point constitue une unité sur laquelle est mesuré le phénomène pour lequel on cherche les composantes d'échelles; on l'appelle unité d'observation ou plus simplement une observation. Elle est repérée par ses coordonnées sur une carte. Le semis des observations doit être régulier, couvrir la totalité de l'aire d'étude, laquelle doit présenter une forme proche du carré. Plus on s'éloigne de ces conditions assez restrictives, plus il est difficile d'interpréter les surfaces comme des composantes d'échelles; par exemple, il est déconseillé de pratiquer ce type d'analyse sur le Togo, le Chili ou le Vietnam.

### 2.2 Les attributs géographiques

Ce sont des propriétés mesurables des observations propres à donner une représentation exhaustive du phénomène étudié. Pour mesurer ces propriétés, il faut recourir à une échelle de mesure choisie parmi les quatre niveaux d'échelles possibles. La plus rudimentaire, l'échelle nominale, représente un phénomène sous forme de modalités; les classifications régionales en constituent un excellent exemple. Les modalités employées pour construire une échelle nominale doivent être exhaustives, c'est-à-dire couvrir l'ensemble des situations tout en s'excluant

mutuellement: une observation ne peut posséder qu'une modalité à la fois. Une échelle binaire est un cas particulier à deux modalités seulement. Une échelle ordinale est établie lorsque toutes les modalités sont représentées par leur rang dans l'ordre croissant ou décroissant de leur valeur. La différence quantitative entre observations reste inconnue. La représentation numérique des rangs fait appel à l'ensemble des nombres entiers privés de zéro. L'échelle d'intervalles, qui ne comporte pas de zéro naturel, permet de calculer une valeur numérique exprimant la différence entre deux observations. Mais en général, on préfère utiliser des échelles de rapport possédant une vraie valeur zéro et rapportant les différences entre observations à une même unité. La plupart des échelles d'intervalle ou de rapport sont continues, c'est-à-dire qu'elles prennent n'importe quelle valeur à l'intérieur de certaines limites. A l'inverse, les échelles discrètes n'ont que des valeurs entières. Une fois adoptée une échelle et après avoir procédé à la mesure proprement dite, il est possible de construire la matrice d'information tendancielle.

### 2.3. La matrice d'information tendancielle

Toute matrice d'information tendancielle est un tableau composé de  $n$  lignes et de quatre colonnes. Les deux premières sont les coordonnées géographiques de chaque point d'observation sur une carte (mesurées en millimètres à partir du coin inférieur gauche). Les abscisses sont désignées par  $X$  et les ordonnées par  $Y$ . La troisième colonne,  $Z$ , est une mesure du phénomène à l'étude utilisant l'une quelconque des quatre échelles précédentes; en pratique, on utilise surtout l'échelle binaire, d'intervalle ou de rapport. Enfin, la quatrième colonne contient l'identifiant du point (un code ou un nom en clair).

2.4. Quelques exemples de matrice d'information tendancielle

Extraites de quelques annuaires statistiques brésiliens, les matrices qui suivent faciliteront la compréhension ainsi que la pratique des trois méthodes composant le système SURF.

2.4.1. La matrice BAHDEN (fig. 2.1)

Les 26 observations composant cette matrice sont les micro-régions homogènes de l'état de Bahia (fig. 2.2). Il s'agit d'agréats de municipios, unité territoriale de base, réalisés en 1968 par l'Institut Brésilien de Géographie et de Statistique. Z étant la densité de population en 1980 (nombre d'habitants par kilomètre carré), il paraît préférable d'affecter à X et Y les coordonnées des sièges des municipios les plus denses, plutôt que celles du centre géométrique par exemple. Cela revient en fait à considérer que la valeur de la densité de population est directement liée à la présence d'une agglomération, ce qui dans l'état de Bahia n'est pas une hypothèse trop forte vue la taille moyenne, assez étendue, des municipios. C'est cette matrice qui a permis de réaliser le premier exemple présenté plus haut.

X	Y	Z	NOM
3.4	13.0	2.62	BARREIRAS
5.0	10.4	3.88	STA. MARIA DA VIT.
7.9	15.9	3.59	XIQUE-XIQUE
6.6	10.7	7.23	BOM JESUS DA LAPA
9.6	14.9	12.60	IRECE
8.0	11.2	9.14	MACAUBAS
10.0	8.8	13.48	BRUMADO
12.9	16.7	11.99	SEN. DO BOMFIM
12.3	15.2	16.82	JACOBINA
12.3	18.9	7.56	JUAZEIRO
15.2	16.7	11.53	EUCLIDES DA CUNHA
15.3	14.2	27.63	SERRINHA
15.4	12.9	43.95	FEIRA DE SANTANA
13.2	9.6	25.60	JEQUIE
11.7	7.5	23.49	VIT. DA CONQUISTA
12.9	6.6	13.34	ITAPETINGA
16.8	19.1	13.40	PAULO ALFONSO
16.5	13.2	34.18	ALAGOINHAS
17.2	13.6	14.39	ENTRE RIOS
16.4	11.4	538.28	SALVADOR
15.9	12.3	85.59	SANTO AMARO
15.2	10.6	31.35	VALENCA
13.3	7.0	21.55	IGUAI
14.8	7.6	48.37	ITABUNA
14.4	2.9	18.37	ITAMARAJU
15.3	4.6	13.78	STA. CRUZ CABRALIA

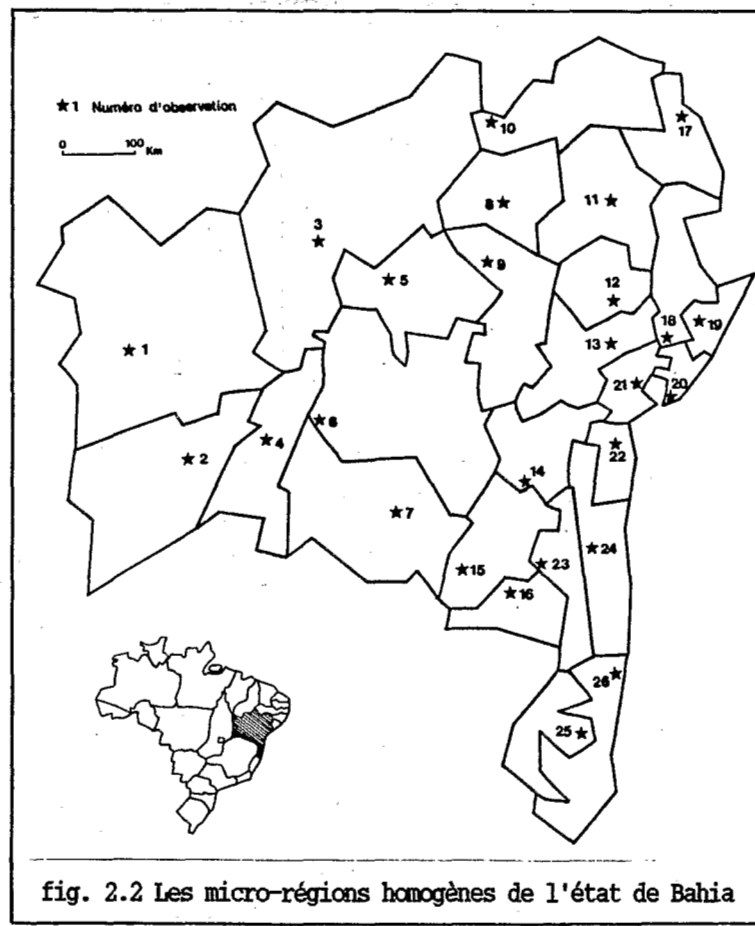
fig. 2.1 La matrice BAHDEN

2.4.2. La matrice SOJA7080 (fig. 2.3)

D'une taille plus imposante que la précédente, cette matrice résulte de la sélection des micro-régions homogènes qui, en 1980, avaient au moins 1000 hectares plantés en soja (fig. 2.4). Il s'agit de la majeure partie des états du Rio Grande do Sul, Santa Catarina, Parana, Mato Grosso do Sul et Sao Paulo, mais aussi d'une part importante du Minas Gerais et Goias. X et Y sont les coordonnées du siège du municipio qui, dans chaque micro-région, a la plus grande superficie en soja. Pour chacune de ces dernières, on mesure la progression du soja par la différence entre la surface (en hectares) qu'il occupe en 1980 et celle qu'il occupait en 1970; ces valeurs forment les coordonnées Z du tableau.

2.4.3. La matrice SOJA12 (fig. 2.5)

Contrairement aux deux précédentes matrices qui avaient une colonne Z mesurée sur une échelle de rapports, celle-ci adopte une échelle binaire. En effet, il s'agit ici d'avoir une idée de la période (1970-1975 ou 1975-1980) durant laquelle la surface en soja s'est le plus accrue. Z prend la valeur 1 si la différence des superficies entre 1970 et 1975 est supérieure à celle de 1975-1980 et 2 si c'est l'inverse. Ce



X	Y	Z	NOM			
31.40	20.20	2610	RS308	33.90	63.90	4386 SP225
29.50	21.10	-3405	RS309	36.50	63.40	1253 SP226
27.50	22.20	-1608	RS313	39.60	62.10	4577 SP227
23.00	22.60	35972	RS314	36.60	59.40	653 SP233
34.00	20.20	2003	RS310	38.00	69.10	1240 SP234
23.70	19.10	71586	RS315	42.30	62.90	52490 SP228
27.40	13.80	94961	RS317	43.30	61.00	79326 SP229
23.30	9.80	16140	RS319	42.30	58.60	22972 SP236
21.20	18.50	52587	RS320	44.50	59.60	19025 SP237
29.20	23.10	1842	RS311	42.80	56.10	6049 SP242
27.60	26.20	37399	RS312	46.10	52.70	917 SP248
18.60	27.40	237873	RS323	45.00	61.60	8304 SP230
18.50	29.70	112811	RS324	44.70	61.00	13957 SP238
24.30	29.60	143601	RS325	47.80	54.90	17999 SP243
26.60	26.70	259330	RS326	48.20	56.00	1387 SP244
20.30	26.70	97540	RS327	33.50	58.80	2471 SP231
22.30	29.10	181137	RS328	33.10	57.30	1168 SP240
24.20	26.60	22662	RS329	31.70	54.70	11410 SP250
24.20	25.50	66917	RS330	34.40	51.80	117225 SP251
29.20	27.80	27440	RS331	35.70	51.00	19602 SP252
20.50	20.90	74538	RS316	45.70	49.10	13662 SP253
11.70	25.30	278550	RS321	40.00	46.50	12412 SP260
21.20	25.90	350254	RS322	50.40	71.30	17415 MG160
34.00	29.50	4157	SC303	58.60	71.30	1128 MG166
30.40	31.40	41337	SC304	45.30	69.20	7319 MG171
27.30	32.20	51038	SC305	50.90	67.60	21318 MG172
26.10	35.10	210992	SC306	54.80	70.30	2064 MG173
34.80	35.80	30432	SC307	40.70	70.50	20027 MG170
38.00	38.60	1275	PA268	40.70	64.60	28769 MG177
35.30	40.20	18659	PA272	43.10	64.10	29486 MG178
34.70	41.60	82414	PA273	45.10	67.90	9373 MG179
36.00	46.00	12476	PA274	52.90	57.00	2131 MG190
33.70	38.20	4985	PA275	54.40	61.90	1402 MG191
33.50	40.60	20135	PA276	58.70	67.00	1386 MG181
26.20	37.70	40212	PA291	44.00	80.60	2188 GO355
32.70	41.80	19590	PA277	24.60	74.80	8239 GO356
25.40	46.30	268138	PA286	38.10	83.10	827 GO354
28.00	43.10	18611	PA287	31.30	73.90	50226 GO357
21.10	42.20	596243	PA288	37.60	75.30	27397 GO358
24.60	36.60	144526	PA289	43.90	72.30	1341 GO359
29.20	40.30	86168	PA290	36.40	72.40	111296 GO360
31.50	50.80	87556	PA279	14.30	56.40	169645 MS342
31.80	49.20	30034	PA280	18.20	65.70	60644 MS339
30.40	49.60	118010	PA281	28.10	67.80	33300 MS340
27.90	48.90	76798	PA282	12.30	52.50	312880 MS344
26.30	49.20	3893	PA283	11.20	61.00	3381 MS338
27.60	47.10	55808	PA284	8.50	58.30	12732 MS341
25.10	48.20	-5877	PA285			

fig. 2.3 La matrice SOJA7080

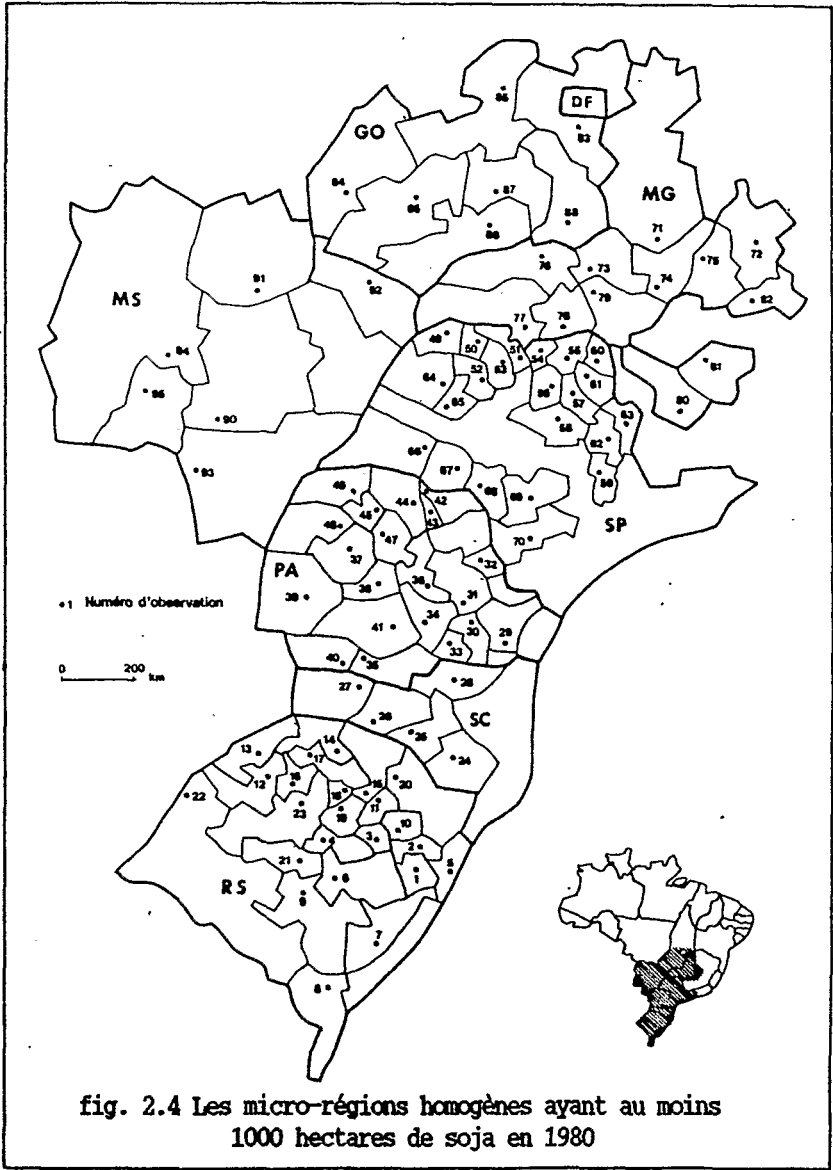


fig. 2.4 Les micro-régions homogènes ayant au moins 1000 hectares de soja en 1980

codage présente l'avantage de ne pas être lié à la taille des micro-régions; à la suite de l'analyse quantitative réalisée sur la matrice SOJA7080, il est intéressant de procéder à une analyse plus qualitative de l'évolution des superficies plantées de soja. Les observations sont les mêmes que pour SOJA7080 et les coordonnées mesurées de la même manière.

## 2.5. Exercice

A. Choisir un espace géographique tel qu'il ait une forme approximativement carrée et au plus une centaine d'unités spatiales. Dresser le fond de carte correspondant en précisant le point le plus représentatif de chaque unité (chef-lieu, centre géométrique...).

B. Etudier un phénomène pouvant, a priori (ce sera votre hypothèse), être décomposé en composantes d'échelles. Relever dans un annuaire statistique les valeurs d'une variable représentant ce phénomène et réaliser, de manière classique, la carte thématique les figurant.

C. Dresser sur une grande feuille de papier la matrice d'information tendancielle, c'est-à-dire un tableau comprenant, en ligne, les unités spatiales, et les quatre colonnes suivantes:

X: coordonnée du point représentant l'unité sur l'axe des abscisses à partir du coin inférieur gauche de la carte.

Y: coordonnée sur l'axe des ordonnées.

Z: valeur de la variable choisie

NOM: nom de l'unité (40 caractères au maximum)

X	Y	Z	NOM
31.40	20.20	1	RS308
29.50	21.10	1	RS309
27.50	22.20	1	RS313
23.00	22.60	1	RS314
34.00	20.20	1	RS310
23.70	19.10	1	RS315
27.40	13.80	1	RS317
23.30	9.80	2	RS319
21.20	18.50	2	RS320
29.20	23.10	1	RS311
27.60	26.20	1	RS312
18.60	27.40	1	RS323
18.50	29.70	1	RS324
24.30	29.60	1	RS325
26.60	26.70	1	RS326
20.30	26.70	1	RS327
22.30	29.10	1	RS328
24.20	26.60	1	RS329
24.20	25.50	1	RS330
29.20	27.80	2	RS331
20.50	20.90	1	RS316
11.70	25.30	1	RS321
21.20	25.90	1	RS322
34.00	29.50	2	SC303
30.40	31.40	2	SC304
27.30	32.20	1	SC305
26.10	35.10	1	SC306
34.80	35.80	2	SC307
38.00	38.60	2	PA268
35.30	40.20	2	PA272
34.70	41.60	1	PA273
36.00	46.00	2	PA274
33.70	38.20	2	PA275
33.50	40.60	2	PA276
26.20	37.70	2	PA291
32.70	41.80	2	PA277
25.40	46.30	1	PA286
28.00	43.10	2	PA287
21.10	42.20	1	PA288
24.60	36.60	1	PA289
29.20	40.30	1	PA290
31.50	50.80	1	PA279
31.80	49.20	1	PA280
30.40	49.60	1	PA281
27.90	48.90	1	PA282
26.30	49.20	1	PA283
27.60	47.10	1	PA284
25.10	48.20	1	PA285
33.90	63.90	1	SP225
36.50	63.40	2	SP226
39.60	62.10	2	SP227
36.60	59.40	1	SP233
38.00	69.10	2	SP234
42.30	62.90	1	SP228
43.30	61.00	1	SP229
42.30	58.60	1	SP236
44.50	59.60	1	SP237
42.80	56.10	2	SP242
46.10	52.70	1	SP248
45.00	61.60	1	SP230
44.70	61.00	2	SP238
47.80	54.90	1	SP243
48.20	56.00	2	SP244
33.50	58.80	1	SP231
33.10	57.30	1	SP240
31.70	54.70	1	SP250
34.40	51.80	1	SP251
35.70	51.00	1	SP252
45.70	49.10	1	SP253
40.00	46.50	1	SP260
50.40	71.30	2	MG160
58.60	71.30	2	MG166
45.30	69.20	1	MG171
50.90	67.60	2	MG172
54.80	70.30	2	MG173
40.70	70.50	1	MG170
40.70	64.60	2	MG177
43.10	64.10	2	MG178
45.10	67.90	1	MG179
52.90	57.00	2	MG190
54.40	61.90	2	MG191
58.70	67.00	2	MG181
44.00	80.60	2	G0355
24.60	74.80	2	G0356
38.10	83.10	1	G0354
31.30	73.90	2	G0357
37.60	75.30	2	G0358
43.90	72.30	2	G0359
36.40	72.40	2	G0360
14.30	56.40	2	MS342
18.20	65.70	2	MS339
28.10	67.80	2	MS340
12.30	52.50	2	MS344
11.20	61.00	2	MS338
8.50	58.30	2	MS341

fig. 2.5 La matrice SOJA12



### 3. Trois méthodes d'analyse en surfaces de tendances.

Les trois méthodes proposées par SURF, bien qu'ayant en commun la recherche d'une tendance spatiale, diffèrent sur deux points essentiels qui doivent être éclaircis avant toute présentation technique.

Une première différence résulte de l'échelle de mesure requise pour exprimer les valeurs de la variable Z. WRIGLEY impose une échelle binaire alors que LEMAY et MATHER exigent une échelle d'intervalle ou de rapport.

Le second point est encore plus important : il s'agit de la méthode de détermination des tendances proprement dite. MATHER et WRIGLEY utilisent la régression polynomiale multiple où X et Y sont les variables exogènes et Z la variable endogène. LEMAY emploie une technique extrêmement différente : c'est une régression linéaire simple entre Z, la variable endogène, et D, la distance calculée entre chaque observation et un point de référence optimal dont la position change à chaque ajustement. Ainsi, les tendances sont cumulatives pour MATHER et WRIGLEY (on trouve la forme de la tendance précédente dans la nouvelle) alors qu'elles sont soustractives pour LEMAY (la nouvelle tendance n'existe qu'une fois éliminée l'influence de la précédente).

#### 3.1. La méthode MATHER

Ce modèle de base de l'analyse en surfaces de tendances postule une relation fonctionnelle entre les valeurs prises par un phénomène (Z) mesuré à l'aide d'une échelle d'intervalle ou de rapport sur n observations localisées sur une carte et leurs coordonnées (X et Y) ; cela est formulé par l'équation suivante :

$$z_i = f(x_i, y_i) + e_i$$

(avec i variant de 1 à n,  
le nombre d'observations)

En un point correspondant à une observation i, la valeur  $z_i$  peut être estimée de

manière approchée par la fonction f. Le résidu  $e_i$ , différence entre la valeur observée et la valeur estimée, reste inexpliqué par la fonction. Le problème revient donc à trouver les coefficients qui, une fois multipliés par les valeurs des coordonnées, produiront l'estimation de  $z_i$  (noté  $\hat{z}_i$ ) c'est-à-dire

$$z_i = \text{cte} + a_1 x_i + a_2 y_i + e_i$$

ou bien encore

$$z_i = \text{cte} + \hat{z}_i + e_i$$

La régression linéaire multiple est une solution adoptant les critères des moindres carrés. L'exposé de cette méthode figure dans tous les ouvrages de statistique ; nous ne nous y attarderons donc pas.

A l'aide des coefficients ( $a_1$  et  $a_2$ ) de la régression, l'estimation de Z en chaque point de la surface rend possible le tracé d'une carte en isolignes reliant les points de même valeur estimée, selon un espacement donné. Pour mieux lire cette carte, on aura toujours intérêt à la superposer à celle localisant les observations.

C'est à ce moment de l'analyse que trois questions doivent être posées. En premier lieu, la surface ajustée, représentant un plan, a-t-elle une certaine valeur ? Une réponse peut être apportée par l'examen du taux de variation expliquée, égal à 0% dans le cas où la fonction ne permet absolument pas de retrouver la variation de Z dans l'espace, et 100% lorsqu'il rend parfaitement compte du phénomène. Ensuite, les valeurs estimées ont-elles la même qualité en chaque point de la carte ? N'existe-t-il pas des zones de sur- ou sous-estimation systématique formant des concentrations de résidus positifs ou négatifs ? La carte des résidus est un moyen pratique de s'en rendre compte. Dans l'affirmative, il faudrait ajuster un autre modèle, par exemple en transformant les valeurs de Z de manière à éviter ce biais systématique. Enfin, si le plan n'explique qu'une petite part de la variation totale, n'est-on pas fondé à chercher une fonction de X et de Y qui aboutisse à de



meilleures estimations ? En pratique, et compte tenu de l'hypothèse stipulant l'existence de plusieurs composantes d'échelles, on préfère un modèle de régression plus complexe mais mieux adapté.

### 3.1.1. Le modèle polynomial

Mather utilise dans son programme le modèle polynomial à un degré (on dit aussi un ordre) variant de 1 à 3. La forme la plus simple de ce modèle de surface de tendance correspond à l'ordre 1 (surface plane) et ne diffère nullement de la régression multiple :

$$z_i = cte + a_1x_i + a_2y_i + e_i$$

A l'ordre 2 (surface quadratique), ce sont 5 coefficients et une constante qu'il faut estimer

$$z_i = cte + a_1x_i + a_2y_i + a_3x_i^2 + a_4x_iy_i + a_5y_i^2 + e_i$$

Enfin, à l'ordre 3 (surface cubique), on obtient une équation avec neuf coefficients et une constante. Théoriquement, rien n'interdit de procéder à des ajustements d'ordre supérieur à 3 (pour peu que le nombre d'observations soit lui-même supérieur au nombre de coefficients). Cependant, plusieurs limitations incitent à ne pas dépasser cette valeur. Tout d'abord, la précision des micro-ordinateurs ne permet pas d'élever indéfiniment la puissance d'un nombre, même en double précision comme le permet le langage FORTRAN. Ensuite, les temps de calcul deviennent vite prohibitifs en raison du grand nombre de coefficients à estimer. Enfin, l'interprétation des résultats apparaît de plus en plus malaisée car les surfaces sont contingentes de variations très locales, ce qui n'est pas, bien au contraire, la finalité de la méthode.

Ainsi adapté à l'analyse en surfaces de tendances, le modèle polynomial aboutit au tracé de trois surfaces rendant compte de mieux en mieux des données (au sens des moindres carrés). Pour chaque surface, on dispose du taux de variation qu'elle explique, de son équation et de la carte représentative en isolignes. Pour compléter l'analyse, la carte des résidus est également très utile.

### 3.1.2. Principes d'interprétation

Les différentes surfaces doivent pouvoir être considérées comme des composantes d'échelle et non pas seulement comme les produits d'ajustements selon des critères géométriques n'ayant rien à voir avec la géographie. C'est bien là que réside la

principale difficulté de la méthode; il faut reconnaître qu'il n'existe pas de règle absolue en ce domaine, l'interprétation se limitant souvent à une description assez stérile, forme de paraphrase plus ou moins habile des sorties d'ordinateur. Deux questions importantes doivent trouver une réponse : 1/ quel est le nombre d'ajustements à interpréter? et 2/ ces ajustements correspondent-ils réellement à autant de composantes d'échelle (générale, régionale et locale)? L'expérience apporte deux éléments de réponse partiels. Tout d'abord, plus les taux de variation expliquée sont dans un rapport élevé, plus il y a de chances pour que les surfaces expriment des composantes d'échelle différentes. En général, les taux diminuent avec l'accroissement du nombre d'observations. Ce sont les proportions qui sont intéressantes. En second lieu, la carte de tendance est un complément indispensable; un accroissement du taux de variance peut apparaître sans que la forme change de manière

significative (dans ce cas, la nouvelle surface n'exprimera pas une nouvelle composante d'échelle). C'est pourquoi il est nécessaire d'examiner sérieusement les cartes en isolignes successives en s'efforçant de lire les grandes lignes plus que les détails. En résumé, c'est la conjonction du résultat de l'examen de ces deux indicateurs (rapports élevés des taux de variation expliquée et différences nettes entre les cartes) qui conduit à accepter une nouvelle tendance exprimant une nouvelle composante d'échelle. De nombreux tests statistiques devraient attribuer à cette méthode plus de rigueur mathématique; mais, élaborés dans le cadre assez contraignant du modèle de régression, ils ont une valeur décisionnelle uniquement lorsque toutes les hypothèses concernant notamment la distribution des résidus sont vérifiées, ce qui n'est qu'assez rarement le cas des données géographiques (cette voie mérite quand-même d'être explorée, en particulier lorsque la recherche a une fin prévisionnelle).

### 3.1.3. Utilisation du programme

A l'appel du progiciel SURF (voir les modalités pratiques au chapitre 4) apparaît pendant quelques secondes l'écran d'accueil (fig. 3.1), remplacé ensuite par le premier menu (fig. 3.2). L'ensemble des menus s'affichant successivement à l'écran est une aide apportée à l'utilisateur lui permettant d'une part de choisir sa méthode d'analyse et, d'autre part, de préparer les données en fonction de ce qu'en attendent les programmes de traitement. Pour choisir la méthode MATHER, il suffit d'appuyer sur la touche "2" et de valider ce choix par la touche "entrée" (fig. 3.2).

Institut Français de Recherche Scientifique  
pour le Développement en Coopération  
ORSTOM

Philippe WANIEZ Yann LE GAUFFEY

```
SSSS U U RRRR FFFFFF
S S U U R R F
S U U R R F
SSSS U U RRRR FFFF
S U U R R F
S S U U R R F
SSSS UUUUU R R F
```

ANALYSE EN SURFACES DE TENDANCES  
OU DE PROBABILITES

fig. 3.1 L'écran d'accueil de SURF

Choisissez une option

- 1 : Surfaces de tendances, methode Lemay
- 2 : Surfaces de tendances, methode Mather
- 3 : Surfaces de probabilités, methode Wrigley
- 4 : Saisie d'une matrice de données
- 0 : Fin de travail

Votre choix? 2

fig. 3.2 Le menu d'options de SURF

Le programme demande alors le nom du fichier contenant les données (fig. 3.3); ce fichier aura préalablement été créé à l'aide de l'option 4 du premier menu (selon des modalités exposées au chapitre 4).

Nom du fichier de données (Arrêt : tapez 0)? a:soja7080.dat

fig. 3.3 Le choix d'un fichier de données

Enfin, SURF offre la possibilité de tracer des cartes en isolignes sur un espace plus ou moins étendu. Par défaut, les coordonnées minimales et maximales de la carte dépendent des minima et maxima des coordonnées des observations (xmin, xmax et ymin, ymax); ce calcul est réalisé si l'opérateur répond -1 à la question "minimum sur l'axe des abscisses?" (fig 3.4). Mais on peut vouloir tracer cette carte sur un espace plus grand ou plus petit que celui à l'intérieur duquel figurent la totalité des observations (les calculs sont cependant effectués sur la totalité des observations); dans

ce cas, l'usager doit donner successivement le minimum et le maximum sur l'axe des abscisses ainsi que le minimum et le maximum sur l'axe des ordonnées. Le tracé des cartes est alors réalisé dans un carré dont l'origine (en bas, à gauche sur la carte) a pour coordonnées le plus petit des deux minima (xmin ou ymin) et dont la coordonnée maximale est le plus grand des deux maxima (xmax ou ymax).

Coordonnées du cadre de la carte  
(-1 calcul direct de ces valeurs)

minimum sur l'axe des abscisses? -1

fig. 3.4 La délimitation du cadre de la carte de tendance

Viennent ensuite deux menus permettant de préciser le mode de sortie des résultats. Le premier (fig. 3.5) donne le choix entre faire sortir à la fois les calculs et les cartes (option 1) ou bien seulement les calculs (option 2); cette dernière n'est utile que pour procéder à des essais. Le menu suivant (fig. 3.6) sert à déterminer la destination des résultats, soit l'imprimante avec l'option 1 (80 caractères par ligne), soit l'écran avec l'option 2 (25 lignes de 80 caractères), soit un fichier de nom standard SURF.OUT à l'aide de l'option 3. En général, on choisit l'option 1 (dans ce cas, l'imprimante doit être connectée à l'ordinateur); la sortie sur écran permet de faire des essais; la sortie dans un fichier est utile lorsqu'on ne dispose pas d'imprimante : il faut alors imprimer le fichier SURF.OUT sur une autre machine dotée de ce périphérique, sans avoir pour cela à relancer le traitement.

Choisissez ...

- 1 : Sortie des calculs et cartes
- 2 : Sortie des calculs seulement

Votre choix? 1

fig. 3.5 Le choix d'une option de sortie des résultats

```

Sortie des resultats

1 : Imprimante
2 : Ecran
3 : Fichier (surf.out)

Votre choix? 1
  
```

fig. 3.6 Le choix d'une unité de sortie des résultats

Un sixième menu (fig. 3.7) propose la transformation de Z selon quatre méthodes (il est extrêmement simple d'en ajouter d'autres en modifiant le programme SURF.BAS comme cela est expliqué en annexe). Cela est très utile pour obtenir des surfaces s'ajustant mieux aux données. Par exemple, dans le cas de la densité de population dans l'Etat de Bahia (matrice BAHDEN), la transformation logarithmique de Z ( $\text{LOG}(Z)$ ) donne de meilleures estimations. La méthode pour choisir une transformation figure habituellement dans les ouvrages de statistique, après l'exposé de la méthode des moindres carrés. Notons que ces transformations ne modifient que les données en cours de traitement, et non pas le fichier en entrée qui reste inchangé, donc utilisable pour une autre analyse.

```

Transformation de Z
(Methodes Lemay et Mather)

1 : Exp (Z)
2 : Log (Z)
3 : 1 / Z
4 : 1 / Z **2
0 : Pas de transformation

Votre choix? 0
  
```

fig. 3.7 Les options de transformation de Z

Enfin, on demande un titre qui sera imprimé en haut de chaque page de la sortie (fig. 3.8).

A la suite de ce dernier menu commence l'exécution du programme MATHER.

```

Titre du travail

7 superficies en soja : evolution 1970-1980 (hectares)

fig. 3.8 Le choix d'un titre
  
```

### 3.1.4. Présentation des sorties

Les figures 3.9 à 3.19 représentent les sorties du programme du programme MATHER sur les données de la matrice SOJA7080 ayant servi d'exemple au premier chapitre, auquel il faut donc se rapporter pour l'interprétation.

Après l'affichage d'une page de référence où apparaît le sigle "SURF" et indiquant le nom de la méthode choisie et le nombre d'observations traitées, le programme présente les données figurant dans le fichier en entrée. C'est, d'une part, la liste complète de la matrice d'information tendancielle (fig. 3.9) et, d'autre part, le graphique localisant les points de mesure (fig. 3.10) identifiés par les valeurs de la colonne OBS de la matrice, contenant l'identifiant de chaque point.

Viennent ensuite trois groupes de deux pages: chaque groupe correspond à l'un des degrés du polynôme ajusté (1, 2 et 3); les deux pages présentent d'abord les coefficients de l'équation de tendance suivis du taux de variance expliqué, puis le tracé de la carte en isolignes. Les figures 3.11 et 3.12 se rapportent à l'ordre 1, 3.13 et 3.14 à l'ordre 2, 3.15 et 3.16 à l'ordre 3. La page suivante donne un résumé des taux de variation expliqués par chaque ajustement (fig. 3.17) pour avoir une vue plus concise du spectre de ces taux.

Le traitement s'achève par l'étude des résidus, c'est-à-dire l'impression du tableau après trois ajustements, avec notamment la valeur de Z estimé (ZEST), la valeur du résidu (RES) et du résidu standardisé (divisé par l'écart-type résiduel), et cela pour chaque observation (fig. 3.18). Enfin, une carte des résidus (fig. 3.19) permet de chercher des agrégats géographiques de résidus positifs (supérieurs à 0.5 écart-type résiduel) ou négatifs (inférieurs à 0.5 écart-type résiduel). Notons enfin que la sortie des résultats est réduite au strict nécessaire, assez facilement utilisable.

soja superficies recoletees evolution 1970-1980					
OBS	X	Y	Z	NOM	
1	31.40	20.20	2610.00	RS308	
2	29.50	21.10	-3405.00	RS309	
3	27.50	22.20	-1608.00	RS313	
4	23.00	22.60	35972.00	RS314	
5	34.00	20.20	2003.00	RS310	
91	18.20	65.70	60644.00	MS339	
92	28.10	67.80	33300.00	MS340	
93	12.30	52.50	312880.00	MS344	
94	11.20	61.00	3381.00	MS338	
95	8.50	58.30	12732.00	MS341	

fig. 3.9 L'édition de la matrice SOJA7080

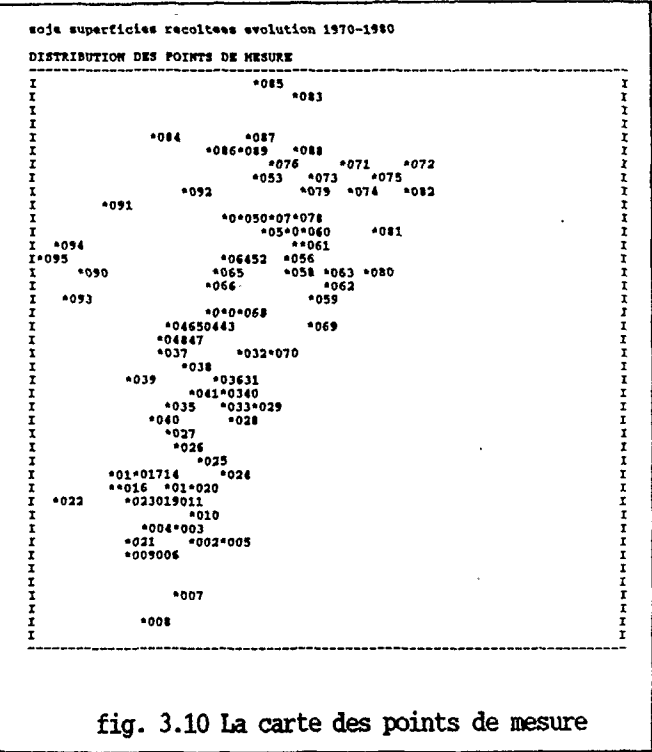
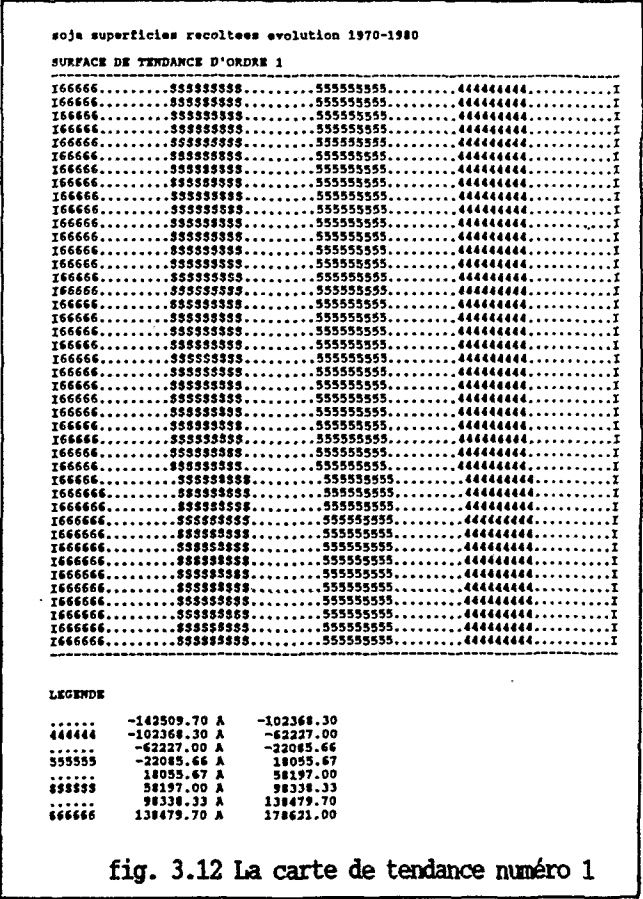


fig. 3.10 La carte des points de mesure

soja superficies recoletees evolution 1970-1980	
SURFACE DE TENDANCE D'ORDRE 1	
COEFFICIENTS DE L'EQUATION	
CTE.	198025.40000
X	-4124.68500
Y	-72.33150
TAUX DE VARIATION EXPLIQUEE = 23.1%	

fig. 3.11 Les paramètres de l'équation de tendance numéro 1





soja superficies recoletees evolution 1970-1980

RESUME POUR 3 AJUSTEMENTS

SURFACE NUMERO	TAUX DE VARIATION EXPLIQUEE	DIFFERENCES
1	23.06	
2	39.63	16.579
3	47.00	7.368

fig. 3.17 Le résumé des résultats de l'analyse de l'évolution du soja

soja superficies recoletees evolution 1970-1980

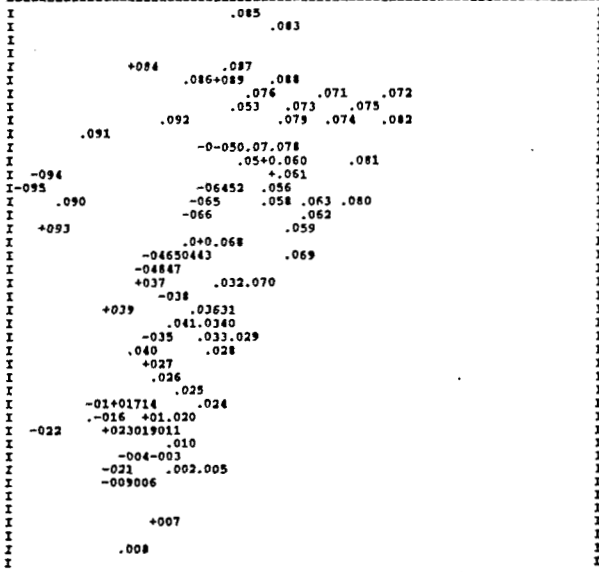
TABLEAU DES RESIDUS APRES 3 AJUSTEMENTS

OBS	X	Y	Z	ZEST	RES	RESSTD NOM
1	31.4	20.2	2610.0	19428.8	-16818.8	-.2 RS308
2	29.5	21.1	-3405.0	23095.3	-26500.3	-.4 RS309
3	27.5	22.2	-1608.0	38423.7	-40031.7	-.6 RS313
4	23.0	22.6	35972.0	93618.8	-57646.8	-.8 RS314
5	34.0	20.2	2003.0	30198.5	-28195.5	-.4 RS310
91	18.2	65.7	60544.0	33171.5	27472.5	.4 MS339
92	28.1	67.8	33300.0	61018.7	-27718.7	-.4 MS340
93	12.3	52.5	312880.0	222513.5	90366.5	1.3 MS344
94	11.2	61.0	3381.0	41564.9	-38183.9	-.6 MS338
95	8.5	58.3	12732.0	69008.4	-56276.4	-.8 MS341

fig. 3.18 le tableau des résidus après trois ajustements

soja superficies recoletees evolution 1970-1980

DISTRIBUTION DES RESIDUS APRES 3 AJUSTEMENTS



LEGENDE  
 - RESIDUS NEGATIFS  
 . RESIDUS MOYENS  
 + RESIDUS POSITIFS

fig. 3.19 La carte des résidus après  
trois ajustements

### 3.2 La méthode WRIGLEY

La méthode d'analyse en surfaces de probabilités, conçue et programmée par Neil Wrigley, est adaptée à des matrices d'information tendancielle où Z est une variable mesurée sur une échelle nominale à 2, 3 ou 4 modalités. En raison des limitations imposées par les micro-ordinateurs, SURF ne propose cette méthode que pour l'échelle binaire (c'est-à-dire à deux modalités).

Sur le plan de la finalité, il n'y a pas de différence sensible entre cette technique et la précédente. Il s'agit d'une régression polynomiale jusqu'au troisième degré, produisant trois cartes de tendance ajustant de mieux en mieux les données initiales pouvant être interprétées comme des composantes d'échelles. Cependant, en raison des caractéristiques mathématiques spécifiques de l'échelle binaire, tout le processus de calcul et d'interprétation se trouve modifié. En effet, puisque  $z_i$  (la valeur des Z observée sur la  $i^{\text{ème}}$  information) ne peut prendre qu'une modalité parmi deux possibles, on doit chercher à estimer le niveau de chance d'observer soit l'une, soit l'autre. Cela revient à considérer le phénomène étudié comme la réalisation d'un certain nombre d'épreuves (autant qu'il y a d'observations), la probabilité d'apparition d'une modalité plutôt qu'une autre variant en fonction des coordonnées géographiques des observations.

#### 3.2.1. Le modèle LOGIT

Si on note  $p1_i$  la valeur de la probabilité pour que  $z_i$  soit égal à la valeur de la première modalité, alors, celle pour que  $z_i$  prenne la valeur de la seconde modalité est  $1 - p1_i$ , soit  $p2_i$ ; connaître l'une des deux probabilités permet de déduire directement l'autre. Par similarité avec le modèle des surfaces de tendances, celui des surfaces de probabilité se formule de la manière suivante:

$$p1 = f(x_i, y_i)$$

avec i variant de 1 jusqu'à n,  
le nombre d'observations

ou bien

$$p2_i = 1 - f(x_i, y_i)$$

Cette approche en forme de probabilités fait que la valeur estimée par la surface est elle-même une probabilité devant par définition être comprise entre 0 inclus et 1 inclus, c'est-à-dire

$$0 \leq \hat{p}1_i \leq 1$$

$$0 \leq \hat{p}2_i \leq 1$$

avec

$$\hat{p}1_i + \hat{p}2_i = 1$$

Malheureusement, l'ajustement par la méthode des moindres carrés rend impossible le respect de cette condition car les valeurs des estimations sont comprises entre -l'infini et +l'infini. Pour pallier cet inconvénient, Wrigley propose d'adopter la fonction logistique, qui a pour équation (fig. 3.20):

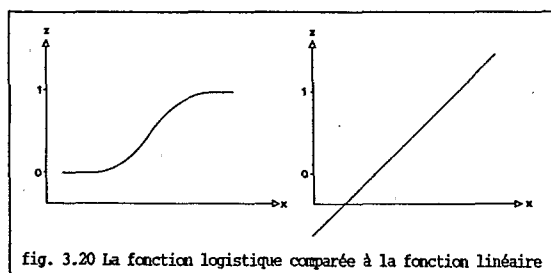


fig. 3.20 La fonction logistique comparée à la fonction linéaire

$$p1_i = \frac{e^{f(x_i, y_i)}}{1 + e^{f(x_i, y_i)}}$$

On peut également écrire ce modèle sous la forme d'une fonction de  $x_i$  et  $y_i$ , c'est-à-dire:

$$\log_e \left( \frac{p1_i}{1 - p1_i} \right) = f(x_i, y_i)$$

La partie gauche de cette équation est connue sous le nom de transformation LOGIT; ses valeurs, notées  $l_i$ , varient entre -l'infini et +l'infini lorsque  $p1$  varie de 0 à 1. En définitive, si on reprend le modèle polynomial, une surface de probabilité d'ordre 1 a pour équation

$$l_i = cte + a_1 x_i + a_2 y_i$$

A l'ordre 2, on obtient:

$$l_i = cte + a_1 x_i + a_2 y_i + a_3 x_i^2 + a_4 x_i y_i + a_5 y_i^2$$

Enfin, à l'ordre 3, ce sont 9 coefficients et une constante qu'il faut estimer à l'aide des observations de la matrice des données. Dans

le cas du modèle des surfaces de tendances, ces paramètres étaient obtenus de manière à rendre maximale la part de variance de  $Z$  expliquée par la surface.

La méthode des moindres carrés étant inadaptée au cas présent, il est nécessaire d'adopter un critère différent de la variance, nommé vraisemblance, qui a pour valeur:

$$V = \pi \sum_{i=1}^{n1} p1_i \pi \sum_{i=n1+1}^n p2_i$$

avec  $n1$  égal au nombre d'observations prenant la modalité 1,  $n$  étant le nombre total d'observations.

Chercher à rendre maximale la vraisemblance revient à estimer les valeurs des paramètres de l'équation de la surface de manière à ce qu'elles aient pu produire, le plus vraisemblablement possible, la répartition spatiale observée des modalités figurant dans la matrice d'information tendancielle. En pratique, compte tenu de l'équation de la fonction logistique, on ne maximise pas directement la vraisemblance, mais son logarithme.

### 3.2.2. Principes d'interprétation

De même qu'en analyse en surfaces de tendances, chaque surface de probabilité doit pouvoir être interprétée comme une composante d'échelle. Les mêmes règles d'interprétation doivent pouvoir être adoptées. En premier lieu, il faut examiner les valeurs d'un indicateur numérique afin d'apprécier le gain de vraisemblance obtenu par chaque surface. Nommé  $RHO^2$ , il se calcule de la manière suivante

$$RHO^2 = 1 - \frac{\log_e V}{\log_e V(C)}$$

Il exprime donc le rapport entre la vraisemblance exprimée par une surface de probabilité d'un ordre donné et celle obtenue par un modèle constant (tel que  $l_i = cte$ ),  $V(C)$ .

Plus  $RHO^2$  a une valeur élevée (variant entre 0 et 1), plus la surface a des chances de représenter une composante d'échelle significative. Il faut également s'intéresser aux valeurs successives de  $RHO^2$  pour chaque surface de probabilité: plus les différences sont importantes, plus le nouvel ajustement risque d'être significatif. Enfin, l'examen des cartes en isolignes, reliant, pour une modalité

donnée, les points d'équi – probabilité, constitue finalement un moyen assez sûr d'interprétation de la valeur de chaque surface. Lorsque sur les cartes apparaissent de nouvelles formes, on aura tout intérêt à s'interroger sur leur signification géographique.

### 3.2.3. Utilisation du programme et présentation des sorties

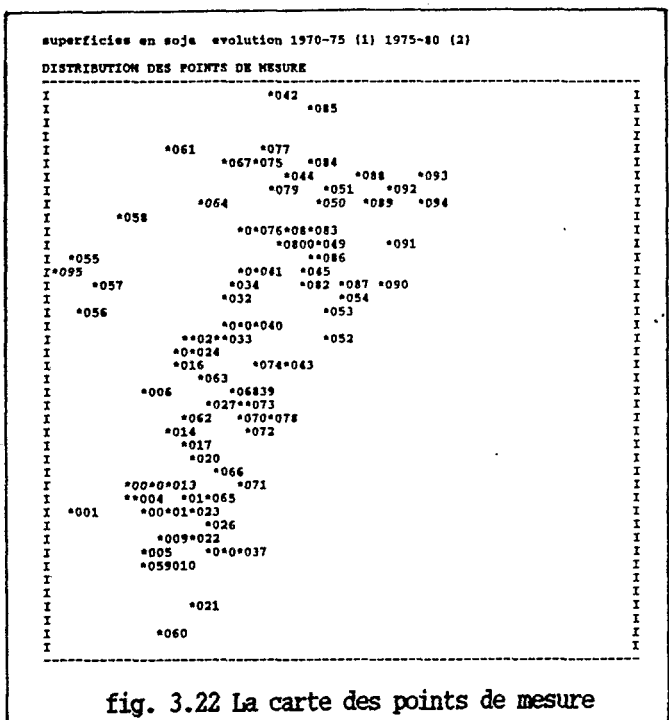
L'utilisation du programme WRIGLEY ne diffère pas sensiblement de celle du programme MATHER présenté en 3.1.3. Il suffit de choisir l'option 3 au premier menu. Le reste du dialogue est identique à l'exception de la transformation de Z (fig. 3.7) inutilisée ici.

Le principe d'impression des résultats est également semblable: 1/ présentation des données initiales (fig. 3.21 et 3.22) 2/ trois groupes de deux pages correspondent à chaque degré du polynôme, d'ordre 1 (fig. 3.24 et 3.25), d'ordre 2 (fig. 3.26 et 3.27) et d'ordre 3 (fig. 3.28 et 3.29): on obtient chaque fois la valeur maximale du logarithme de la vraisemblance puis les coefficients de l'équation de tendance et enfin, le tracé de la carte en isolignes 3/ résumé des ajustements successifs (fig. 3.30): on y trouve pour chaque ajustement a/ la valeur estimée du maximum du logarithme de la vraisemblance (l'ajustement 0 correspond au modèle constant dont l'estimation figure juste avant les calculs de la première surface de probabilité (fig. 3.23), b/ la différence de vraisemblance avec l'ajustement précédent, c/ la valeur de  $RHO^2$  permettant d'apprécier la dissimilarité avec le modèle constant, 4/ le traitement s'achève par l'étude des résidus (fig. 3.31 et 3.32), impression du tableau et tracé de la carte. Notons que toutes ces sorties ne considèrent que la modalité numéro 1 (prenant arbitrairement la valeur 1 pour les calculs), laissant de côté la modalité 2 (prenant dans le processus de calcul la valeur arbitraire 0) qui dans le cas d'une échelle linéaire, est le strict négatif de la modalité numéro 1.

La lecture de la carte de probabilité étant relativement différente de celle des tendances, il est utile de s'y arrêter quelque peu. Prenons en exemple la surface de probabilité du second degré, celle qui ajuste le mieux la distribution observée (fig. 3.27),  $RHO^2$  étant ici plus élevé qu'ailleurs. Cette surface représente la probabilité qu'il y a d'observer la modalité numéro 1, c'est-à-dire un accroissement des surfaces de soja plus grand durant la période 1970-1975 que pendant celle couvrant 1975-1980. Les lignes de probabilités se

superficies en soja evolution 1970-75 (1) 1975-80 (2)				
OBS	X	Y	Z	NOM
1	.117000	.253000	1	RS321
2	.185000	.297000	1	RS324
3	.186000	.274000	1	RS323
4	.203000	.267000	1	RS327
5	.205000	.209000	1	RS316
91	.544000	.619000	2	MG191
92	.548000	.703000	2	MG173
93	.586000	.713000	2	MG166
94	.587000	.670000	2	MG181
95	.085000	.583000	2	MS341
X ET Y ONT ETE MULTIPLIES PAR				.01000

fig. 3.21 L'édition de la matrice SOJA12



superficies en soja evolution 1970-75 (1) 1975-80 (2)	
MAXIMUM DU LOG DE VRAISSEMBLANCE POUR UN MODELE CONSTANT	
MAXIMUM DU LOG DE VRAISSEMBLANCE = -64.95671	

fig. 3.23 Le maximum du logarithme de la vraisemblance pour un modèle constant





pas plus de chance pour que l'accroissement durant la première période dépasse ou soit inférieur à celui de la seconde période. Au-delà de cette ligne d'équi-probabilité, c'est pendant la seconde période que l'accroissement a toutes les chances d'être plus élevé, c'est-à-dire que  $p1_i$  est inférieur à  $p2_i$ .

Cette approche des probabilités est somme toute assez suggestive, une fois débarrassée de son aspect le plus technique.

superficies en soja evolution 1970-75 (1) 1975-80 (2)

RESUME POUR 3 AJUSTEMENTS

AJUSTEMENT NUMERO	MAXIMUM DU LOG DE VRAISSEMBLANCE	DIFFERENCES	D.F.	RHO2
----------------------	--	-------------	------	------

0	-64.957			
		5.975	2	.092
1	-58.981			
		2.058	3	.124
2	-56.924			
		-1.610	4	.099
3	-58.534			

fig. 3.30 Le résumé des résultats de l'analyse des périodes d'évolution du soja

fig. 3.28 Les paramètres de l'équation de tendance numéro 3

superficies en soja evolution 1970-75 (1) 1975-80 (2)

TABLEAU DES RESIDUS APRES 3 AJUSTEMENTS  
MODALITE NUMERO 1

OBS	X	Y	Z	ZEST	RES	RESSTD	NOM
1	.117000	.253000	1	.76017	.23983	.56177	RS321
2	.185000	.297000	1	.74361	-.58720	.58324	RS322
3	.188000	.274000	1	.75686	.24314	.56678	RS323
4	.203000	.267000	1	.76224	.23776	.55850	RS327
5	.205000	.209000	1	.79120	-.20860	.51372	RS316
91	.544000	.619000	2	.40653	-.40653	-.82765	HG191
92	.548000	.703000	2	.29683	-.29683	-.64971	HG173
93	.586000	.713000	2	.27434	-.27434	-.61486	HG166
94	.587000	.670000	2	.32815	-.32815	-.69888	HG181
95	.005000	.583000	2	.48434	-.48434	-.96915	HG341

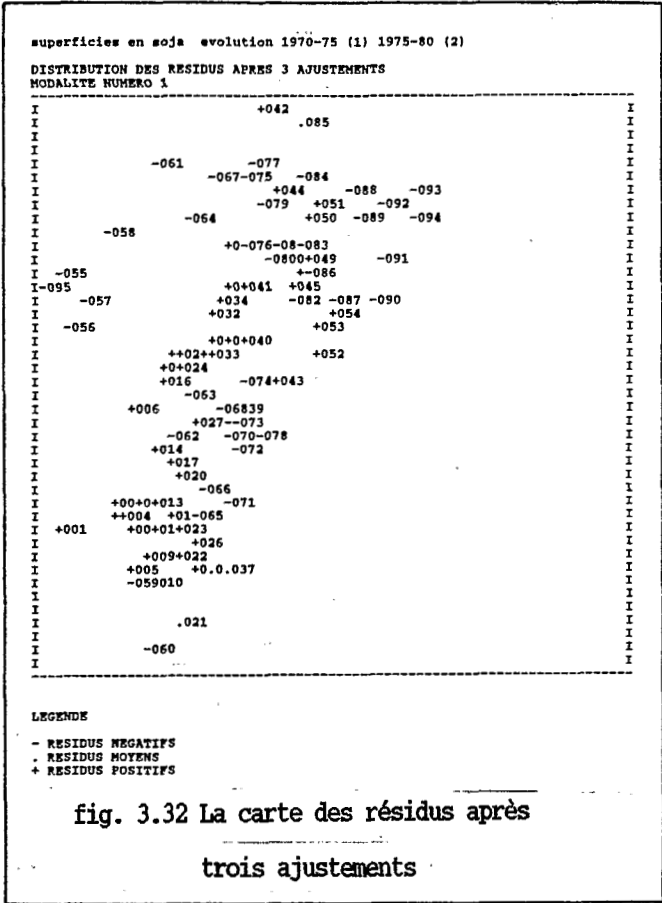
fig. 3.31 Le tableau des résidus après trois ajustements

```

.....      .0000 A      .1000
444444      .1000 A      .2000
.....      .2000 A      .3000
555555      .3000 A      .4000
.....      .4000 A      .5000
888888      .5000 A      .6000
.....      .6000 A      .7000
666666      .7000 A      .8000
.....      .8000 A      .9000

```

fig. 3.29 La carte de tendance numéro 3



### 3.3 La méthode LEMAY

Dans sa conception et dans ses résultats, la méthode conçue et programmée par Guy Lemay est assez différente des deux précédentes. Certes, il s'agit toujours de décomposer un phénomène géographique en tendances de moins en moins générales, mais le chemin pour y arriver apparaît bien plus original et d'un grand potentiel pour la recherche géographique. Le modèle de base, qui, rappelons-le, ne s'applique qu'à une variable  $Z$  mesurée sur une échelle d'intervalles ou de rapports, est formulé par l'équation suivante:

$$z_i = \sum_{k=1}^p \left( a_k + b_k d_{o_k i} + e_i \right)$$

La valeur de la variable  $z$  en un point  $i$  est la somme des valeurs estimées à l'aide de  $p$  fonctions linéaires (en général, trois au maximum) de la distance par rapport à un point optimal  $d_{o_k i}$  différent pour chaque fonction; à cela s'ajoute un résidu  $e_i$  inexpliqué par les fonctions successives.

#### 3.3.1. Réitérer une régression linéaire simple: méthode et interprétation

La méthode de décomposition en tendances est assez simple, mais nécessite un grand volume de calculs influant directement sur la durée d'exécution du programme (qui peut atteindre entre 10 et 15 minutes). En voici la présentation pas à pas.

a. Recherche sur l'ensemble de la surface des coordonnées du point optimal. Il est choisi de manière à rendre maximale la valeur absolue d'un coefficient de corrélation linéaire. Ce coefficient mesure le degré de liaison entre la variable  $Z$  et la variable  $D$  contenant les distances de chaque observation au point optimal. Il faut donc calculer tous les coefficients possibles correspondant à tous les points de l'espace; en pratique, ces calculs sont réalisés avec les coordonnées des noeuds d'une maille rectangulaire placée sur la carte. Le noeud obtenant la valeur absolue la plus élevée est considéré comme le point optimal

b. Tracé de la carte des corrélations. C'est une carte en isolignes joignant les points d'égale corrélation. Elle est donc de nature très différente de celles produites par les méthodes précédentes. Pour faciliter la lecture de cette carte, on peut adopter le point de vue suivant: on a affaire à un gradient d'inaptitude de l'espace à accueillir le phénomène mesuré par la variable  $Z$ . Un point situé dans une zone de corrélations positives est particulièrement inapte puisque les valeurs de  $Z$  tendent à s'accroître en fonction de son éloignement. A l'inverse, lorsqu'un point se trouve dans une zone de corrélations négatives, il se révèle plus apte à voir se dérouler le phénomène puisque les valeurs de  $Z$  diminueront avec la distance.

c. Ajustement d'une fonction linéaire. Une fois trouvé le point optimal, on réalise une régression linéaire simple dans laquelle  $Z$  est la variable endogène (à expliquer) et  $D$ , la variable exogène contenant les distances des observations au point optimal, c'est-à-dire

$$z_i = f(d_{oi}) + e_i$$

Avec cet ajustement sont calculés d'une part, le taux de variance expliquée par la tendance et d'autre part, les résidus formant une nouvelle variable  $E$  telle que

$$e_i = z_i - \hat{z}_i$$

d. Réitération des phases a, b et c. On remplace Z par E de manière à expliquer les résidus par une nouvelle tendance. Afin d'obtenir trois surfaces de tendances, on réitère deux fois. Cette limite de trois ajustements est justifiée de manière pratique en raison du temps

d'exécution du programme qui, rappelons-le, tournait à l'origine sur un gros ordinateur IBM 370; sur micro-ordinateur type IBM PC, on ne peut envisager d'aller au-delà de trois sans aboutir à des temps d'exécution inadmissibles.

Outre l'originalité des cartes de tendance, la méthode LEMAY est extrêmement intéressante: elle rend possible la mise à l'épreuve d'hypothèses sur la forme des relations spatiales. En effet, le programme propose lors de la phase "a" un certain nombre de transformations des distances. A la distance euclidienne usuelle, on substitue alors son inverse, logarithme, exponentielle ou carré. Par exemple, si l'on cherche à prouver l'existence d'une relation du type gravitation, il est utile de transformer la distance par son inverse au carré, c'est-à-dire

$$d_i = \frac{1}{d_1^2}$$

Ainsi, le modèle de base devient

$$z_i = \sum_{k=1}^p \left( a_k + \frac{b_k}{d_{0k}^2} \right)$$

Notons que dans ce cas, l'interprétation de la carte des corrélations se trouve modifiée: une zone de forte corrélation devient une zone d'aptitude élevée.

### 3.3.2. Utilisation du programme et présentation des sorties

Le programme LEMAY s'utilise exactement comme le programme MATHER. Il suffit de choisir l'option numéro 1 après l'affichage du premier menu. Une seule différence apparaît avec le menu proposant la transformation des distances (fig. 3.33).

Le principe d'impression des résultats est également semblable: 1/ Présentation des données initiales (fig 3.34 et 3.35). 2/ Trois groupes de deux pages correspondant à chaque ajustement, le premier expliquant la plus grande part de variance (fig. 3.36 et 3.37), le second (fig. 3.38 et 3.39) et le troisième (fig. 3.40 et 3.41). A chaque fois, le programme

le programme sort le type de l'équation de tendance, le taux de variance expliquée et son cumul avec les précédents ajustements et l'estimation des paramètres; enfin, la carte des corrélations est tracée. 3/ Après le tracé de la troisième surface, le résumé rappelle les coordonnées du point optimal, le taux de variance expliquée et les taux cumulés (fig. 3.42). 4/ Enfin, dans le but de procéder à un examen des résidus, le tableau et la carte des résidus forment un complément extrêmement utile (fig. 3.43 et 3.44).

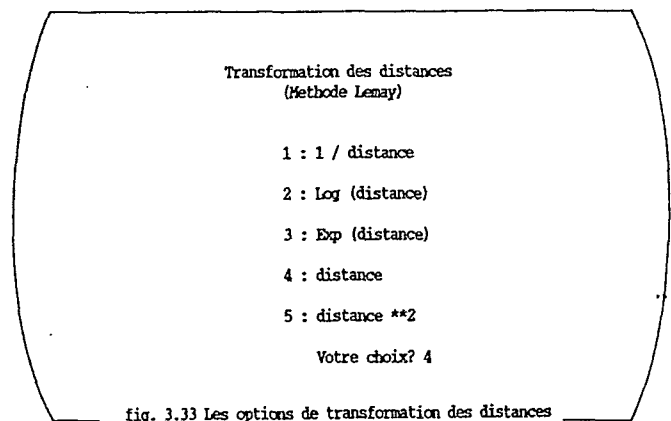
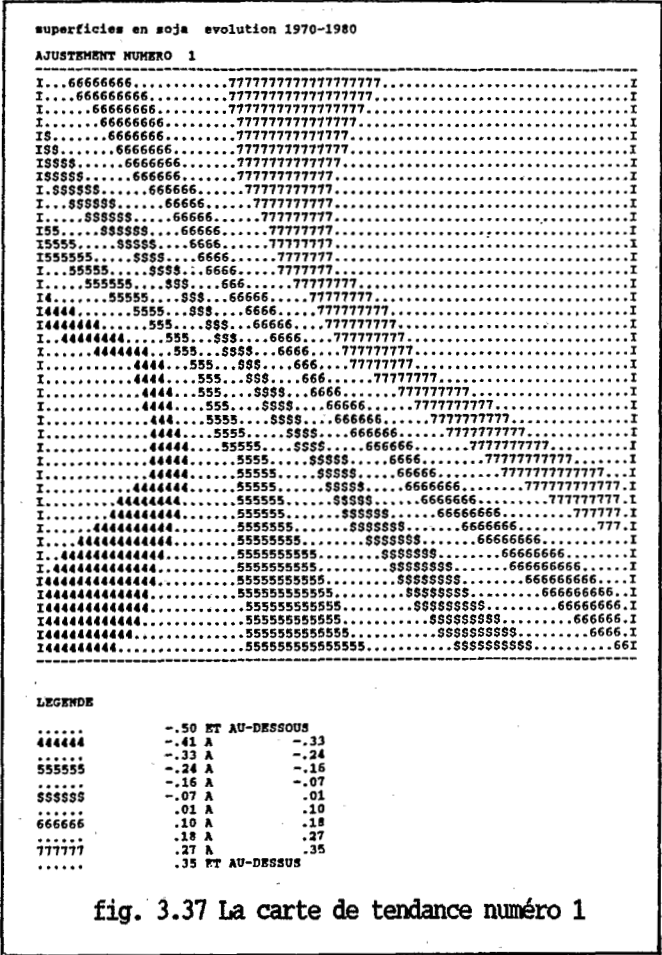
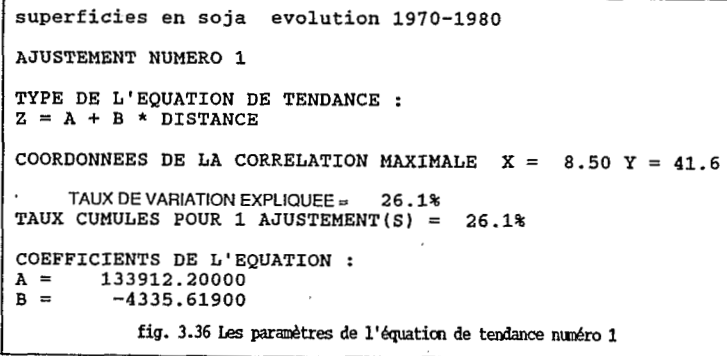
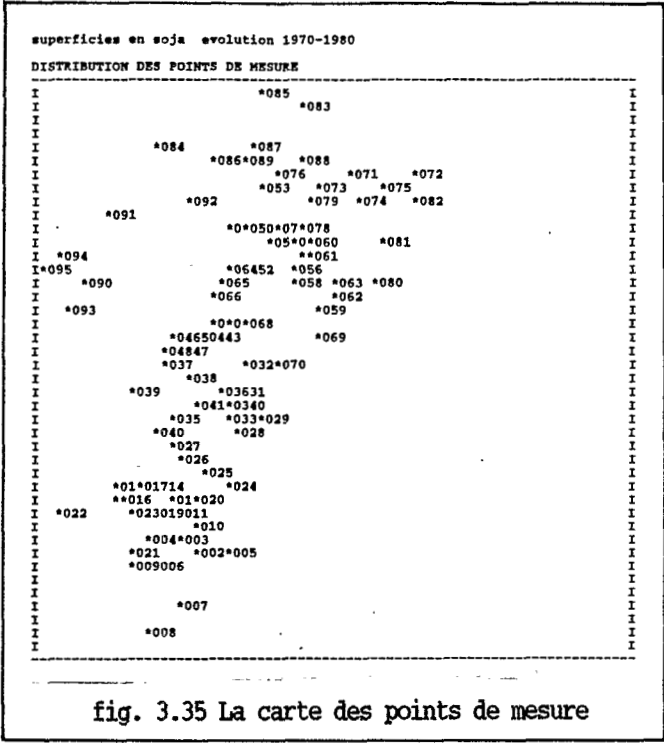


fig. 3.33 Les options de transformation des distances

superficies en soja evolution 1970-1980				
OBS	X	Y	Z	NOM
1	31.40	20.20	2610.00	RS308
2	29.50	21.10	-3405.00	RS309
3	27.50	22.20	-1608.00	RS313
4	23.00	22.60	35972.00	RS314
5	34.00	20.20	2003.00	RS310
91	18.20	65.70	60644.00	MS339
92	28.10	67.80	33300.00	MS340
93	12.30	52.50	312880.00	MS344
94	11.20	61.00	3381.00	MS338
95	8.50	58.30	12732.00	MS341
	33.06	47.86	58197.36	MOYENNES

fig. 3.34 L'édition de la matrice SOJA7080



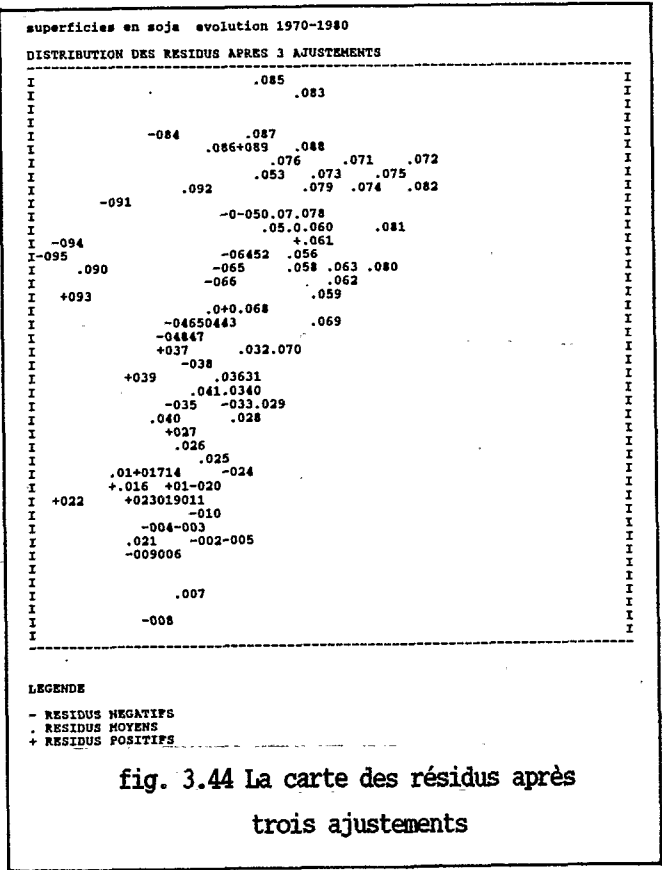


superficies en soja evolution 1970-1980				
RESUME POUR 3 AJUSTEMENTS				
AJUSTEMENT NUMERO	ORIGINES X Y		TAUX DE VARIATION EXPLIQUEE (%)	TAUX DE VARIATION CUMULES (%)
1	8.50	41.66	26.14	26.14
2	43.04	40.27	4.78	30.93
3	20.93	41.66	.76	31.68

fig. 3.42 Le résumé des résultats de l'évolution du soja

superficies en soja evolution 1970-1980							
TABLEAU DES RESIDUS APRES 3 AJUSTEMENTS							
OBS	X	Y	Z	ZEST	RES	RESSTD	NOM
1	31.4	20.2	2610.0	57077.6	-54467.6	-.7	RS308
2	29.5	21.1	-3405.0	66371.8	-69776.8	-.9	RS309
3	27.5	22.2	-1608.0	76749.1	-78357.1	-1.0	RS313
4	23.0	22.6	35972.0	97781.0	-61809.0	-.8	RS314
5	34.0	20.2	2003.0	45782.8	-43779.8	-.6	RS310
91	18.2	65.7	60644.0	110441.8	-49797.8	-.6	MS339
92	28.1	67.8	33300.0	71006.0	-37706.0	-.5	MS340
93	12.3	52.5	312880.0	167097.1	145782.9	1.9	MS344
94	11.2	61.0	3381.0	144058.6	-140677.6	-1.8	MS338
95	8.5	58.3	12732.0	158912.0	-146180.0	-1.9	MS341

fig. 3.43 Le tableau des résidus après trois ajustements



4. Conclusion: comment choisir l'une des trois méthodes ?

Le progiciel SURF propose trois méthodes assez différentes d'analyse en surfaces de tendances, pouvant toutes les trois servir à l'identification de composantes d'échelles. Il apparaît utile de proposer ici quelques critères de choix découlant directement de l'exposé précédent de ces méthodes.

Un premier critère est celui de la nature des données. Pour LEMAY et MATHER, la variable Z doit être mesurée sur une composante d'échelle ou de rapport; WRIGLEY se contente d'une échelle binaire. Si ce dernier cas est celui de vos données, il faut choisir la méthode WRIGLEY.

Un second critère de choix est la présence ou l'absence d'hypothèses sur la nature des relations entre Z et l'espace géographique. Si, par exemple, on émet l'hypothèse d'un modèle gravitaire sous-jacent (étude d'un réseau urbain comme le fait Guy Lemay pour la Champagne et la Picardie), la méthode LEMAY est mieux adaptée. En l'absence d'hypothèse de ce type, il est préférable d'adopter la méthode MATHER, puis, éventuellement selon la nature des résultats obtenus, éprouver un modèle avec la méthode LEMAY.

Le troisième critère, important dans la pratique, concerne le temps d'exécution du programme: MATHER est le plus rapide, suivi de LEMAY et, bien plus loin de WRIGLEY. Les temps d'exécution varient considérablement d'un micro-ordinateur à l'autre, en général de quelques minutes seulement pour MATHER à près d'une demi-heure pour WRIGLEY, pour l'analyse sur les 95 observations du problème "soja".

---

### Eléments de bibliographie

- HAGETT (P.), CHORLEY (R.J.) - 1965 - Trend Surfaces mapping in Geographical research.  
Transaction of British Geographical Institute, n° 37, pp 47-67.
- LE GAUFFEY (Y.), WANIEZ (Ph.) - 1984 - L'analyse en surfaces de tendance.  
Informatique Informations, n°5,  
CIREGE, Université de Paris X-Nanterre, 190 p.
- LEMAY (G.) - 1975 - Méthodes d'analyse chrono-spatiale: les villes de la Champagne  
et de la Picardie.  
Reims, Institut de Géographie, Thèse de 3<sup>e</sup> cycle, 271 p. plus annexes.
- RIDDEL (J.B.) - 1970 - The spatial dynamics of modernization in Sierra-Leone.  
Northwestern University press, 142 p.
- UNWIN (D.) - 1975 - An introduction to trend surface analysis.  
Norwich, CATMOG n° 5, 40 p.
-



## Annexe: le progiciel SURF

Le progiciel SURF n'est pas vendu mais offert avec l'achat du fascicule; seul le prix de la disquette support est payé. Il n'y a aucune garantie contractuelle sur les fichiers qu'elle contient. La duplication est autorisée à condition que les utilisateurs fassent la citation suivante dans leurs publications: "Progiciel SURF - ORSTOM - 1987".

### A.1. Configuration nécessaire

Le système a été testé sur 4 micro-ordinateurs différents, compatibles IBM PC: KAYPRO 16, COMPAQ II, NEXUS 1600, IBM PC XT; que les propriétaires de ces matériels trouvent ici l'expression de nos remerciements. Seuls les temps d'exécution différent, de manière très sensible, le COMPAQ II s'avérant le plus rapide.

La configuration minimale requise est la suivante: 512 Koctets de mémoire RAM, une unité de disquette, un disque magnétique, PC DOS version 2, BASICA version 3 ou GWBASIC ou QUICKBASIC, un éditeur de textes du genre PROFESSIONAL EDITOR (nous avons utilisé la version 1). Une imprimante (80 caractères par ligne et 66 lignes par page) complète utilement cette configuration.

### A.2. La disquette SURF

La disquette de diffusion contient

1/ le système SURF proprement dit (SURF.BAS),

2/ le module exécutable des programmes correspondant à chaque méthode d'analyse (LEMAY.EXE et MATHER.EXE, disquette n°1, WRIGLEY.EXE),

3/ les jeux de données présentés dans cet ouvrage et par les auteurs des programmes (fichiers ayant l'extension .DAT).

### A.3. Installation et appel du progiciel

L'installation de SURF sur votre micro-ordinateur revient simplement à copier tous les fichiers de la disquette de diffusion sur le disque magnétique. Voici le détail des opérations à réaliser:

-insérer la disquette dans l'unité A:, puis entrer la commande

COPY A:\*. \* C:

-retirer la disquette de l'unité A:

L'appel de SURF commencera en entrant la commande BASICA SURF si votre ordinateur est de la marque IBM, GWBASIC SURF si c'est un compatible.

### A.4 . Saisir un fichier de données

A l'aide du progiciel installé de la manière décrite ci-dessus, il est possible de réaliser soi-même les traitements présentés au chapitre 3, ainsi que ceux exposés par les auteurs des méthodes dans leurs ouvrages respectifs. Un excellent exercice consiste à réaliser ces traitements "pour voir comment ça marche !". Au-delà de cette première expérimentation, qu'on souhaite fructueuse, il est intéressant d'analyser son propre jeu de données, par exemple celui mis en forme dans une matrice d'information tendancielle, comme le propose l'exercice du chapitre 2.

Deux cas peuvent se présenter. Dans le premier, vous disposez du PROFESSIONAL EDITOR IBM, vous suivrez donc ce qui est exposé ici; dans le second, vous utilisez un autre éditeur de texte (WORDSTAR, WORDPERFECT, EDLIN...): vous tenterez d'adapter ce qui est dit ici; préalablement à tout appel de SURF, il faudra saisir votre fichier avec cet éditeur.

Après l'appel de SURF et l'affichage de l'écran d'accueil, un menu vous propose diverses options de traitements. L'option 4 offre la possibilité de saisir une matrice de données (fig. A.1); on demande ensuite le nom du fichier à saisir, c'est-à-dire à créer, à étendre ou à corriger. Il est conseillé de choisir un nom rappelant le contenu du fichier (fig. A.2), avec l'extension DAT (ici, la matrice SOJA7080 présentée au chapitre 2). Alors s'exécute PROFESSIONAL EDITOR, commençant par l'apparition de son propre écran d'accueil (fig. A.3) suivi par le menu d'options (fig. A.4). Après l'usage de la touche de fonction 9, apparaît la matrice SOJA7080, déjà saisie. Pour l'utilisation de PROFESSIONAL EDITOR, il est conseillé d'avoir sous la main le manuel de référence IBM.

Choisissez une option

- 1 : Surfaces de tendances, methode Lenay
- 2 : Surfaces de tendances, methode Mather
- 3 : Surfaces de probabilités, methode Wrigley
- 4 : Saisie d'une matrice de données
- 0 : Fin de travail

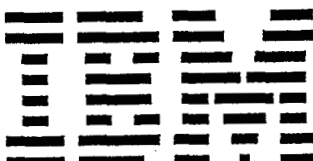
Votre choix? 4

fig. A.1 Le menu d'options de SURF

Nom du fichier de données (Arrêt : tapez 0)? a:soja7080.dat

fig. A.2 Le choix d'un fichier à saisir

The



Personal Computer

Professional Editor

By: Walter J. Paul

Version 1.00 - (C) Copyright IBM Corporation, 1982

fig. A.3 L'écran d'accueil du PROFESSIONAL EDITOR

The IBM Personal Computer - Professional Editor

Primary Options

INPUT FILE — A:SOJA7080.DAT File Size = 3041 Bytes

OUTPUT FILE — A:SOJA7080.DAT

PROFILE FILE — A:DAT.PRF Options set from this profile.

LIST FILES — A:?????????.???

DISPLAY MODE — 3

- 1 = 80 x 25 Monochrome
- 2 = 80 x 25 B & W
- 3 = 80 x 25 Color

DEFAULT DISK — A

Text Buffer Size is 65535 Characters

01 KEYS 2 3 4OPTIONS TABS 6HACROS7FILES 8RESET9 EDIT 0RETURN KB

fig. A.4 Le menu d'options du PROFESSIONAL EDITOR (version 1.00)

La matrice d'information tendancielle doit être saisie de la manière suivante: 1/ chaque observation constitue un enregistrement (fig. A.5), ou, si on préfère, une ligne du fichier. 2/ chaque enregistrement est composé de 4 éléments: la coordonnée sur l'axe des abscisses et sur l'axe des ordonnées (X et Y), la valeur de la variable Z pour laquelle on cherche une tendance spatiale, et enfin, un identifiant pouvant avoir jusqu'à 40 caractères de longueur. Notons que les valeurs décimales sont précédées d'un point ( . ) et que chaque élément est séparé de son successeur par un blanc (le caractère "espace"); il n'est pas nécessaire d'aligner les valeurs des enregistrements comme cela est fait sur la figure A.5, mais cela est plus clair ainsi.

Après avoir sauvé le contenu de la saisie et pressé sur la touche de fonction numéro 10 du premier menu d'option (fig. A.4), le système retourne dans SURF et propose l'exécution d'une des méthodes d'analyse (option 1, 2 ou 3 sur la figure A.1), la saisie ou la correction d'une nouvelle matrice (option numéro 4) ou la fin du travail et le retour au système d'exploitation (option 0). Maintenant, vous pouvez saisir votre propre matrice !

START OF TEXT			
.....	31.40	20.20	2610 RS308
.....	29.50	21.10	-3405 RS309
.....	27.50	22.20	-1608 RS313
.....	23.00	22.60	35972 RS314
.....	34.00	20.20	2003 RS310
.....	23.70	19.10	71586 RS315
.....	27.40	13.80	94961 RS317
.....	23.30	9.80	16140 RS319
.....	21.20	18.50	52587 RS320
.....	29.20	23.10	1842 RS311
.....	27.60	26.20	37399 RS312
.....	18.60	27.40	237873 RS323
.....	18.50	29.70	112811 RS324
.....	24.30	29.60	143601 RS325
.....	26.60	26.70	259330 RS326
.....	20.30	26.70	97540 RS327

Complete file has been read into buffer.

fig. A.5 Le fichier SOJA 7080

### A.5. Ajouter une ou plusieurs options de transformation de Z

Dans le cas des méthodes LEMAY ou MATHER, la transformation de Z est une possibilité intéressante permettant souvent d'obtenir un meilleur ajustement. SURF propose 4 transformations, mais ce nombre n'est pas limitatif tant il est facile d'en ajouter d'autres (pour peu qu'on connaisse quelques rudiments de BASIC).

La figure A.6 présente le fragment concerné du programme SURF (contenu en format caractère dans le fichier SURF.BAS). Pour ajouter une transformation (par exemple, racine carrée de  $z_i$ ), il faut ajouter les lignes suivantes:

```
1995 PRINT "5:RAC(Z)"
1996 PRINT
```

```
2175 IF US$="5" THEN Z=SQR(Z)
```

Ensuite, il faut modifier la ligne 2040 en écrivant:

```
2040 GOTO 2090
```

Brasilia, le 10 Septembre 1987

```
1830 REM
1840 REM TRANSFORMATION DE Z (LEMAY OU MATHER)
1850 REM
1860 REM
1870 CLS
1880 PRINT : PRINT : PRINT : PRINT
1890 PRINT " Transformation de Z"
1900 PRINT " (Methodes Lemay et Mather)"
1910 PRINT : PRINT :
1920 PRINT " 1 : Exp (Z)"
1930 PRINT
1940 PRINT " 2 : Log (Z)"
1950 PRINT
1960 PRINT " 3 : 1 / Z"
1970 PRINT
1980 PRINT " 4 : 1 / Z **2"
1990 PRINT
2000 PRINT " 0 : Pas de transformation"
2010 PRINT
2020 SOUND 150,10
2030 INPUT " Votre choix";US$
2040 IF US$="0" OR US$="1" OR US$="2" OR US$="3" OR US$="4" THEN GOTO 2090
2050 PRINT
2060 PRINT " ERREUR ! Ce choix n'est pas permis"
2070 PLAY "A" : FOR T=1 TO 600 : NEXT T
2080 GOTO 1870
2090 OPEN "O",#2,"FICH2.SUR"
2100 ON ERROR GOTO 2300
2110 FOR I=1 TO 1000
2120 INPUT #1,Z,X,Y,NOMS$
2130 IF US$="1" THEN Z=EXP(Z)
2140 IF US$="2" THEN Z=LOG(Z)
2150 IF US$="3" THEN Z=1/Z
2160 IF US$="4" THEN Z=1/(Z**2)
2170 IF US$="0" THEN Z=Z
2180 PRINT #2,Z,X,Y,NOMS$
2190 IF EOF(1) THEN GOTO 2210
2200 NEXT I
2210 CLOSE #1 : CLOSE #2 :
2220 OPEN "O",#1,"FICH1.SUR" : OPEN "I",#2,"FICH2.SUR"
2230 FOR I=1 TO 1000
2240 INPUT #2,Z,X,Y,NOMS$
2250 PRINT #1,Z,X,Y,NOMS$
2260 IF EOF(2) THEN GOTO 2280
2270 NEXT I
2280 CLOSE #1 : CLOSE #2 : OPEN "I",#1,"FICH1.SUR"
2290 GOTO 2380
2300 PRINT
2310 PRINT " ERREUR ! conditions de la"
2320 PRINT
2330 PRINT " transformation non remplies"
2340 PLAY "A" : FOR T=1 TO 600 : NEXT T
2350 CLOSE #1 : CLOSE #2 :
2360 OPEN "I",#1,"FICH1.SUR"
2370 RESUME 340
2380 IF CS<>"I" THEN GOTO 2660
```

fig. A.6 Un fragment du programme SURF.BAS

## Table des matières

0. SURF, un progiciel d'analyse en surfaces de tendances	3
1. L'analyse en composantes d'échelles et la cartographie des tendances	5
1.1. Un exemple élémentaire: les densités de population dans l'état de Bahia	5
1.2. Un exemple plus complexe: l'expansion de la culture du soja au Brésil	8
2. Elaborer une matrice d'information tendancielle	11
2.1. Les unités spatiales	11
2.2. Les attributs géographiques	11
2.3. La matrice d'information tendancielle	11
2.4. Quelques exemples de matrice d'information tendancielle	12
2.4.1. La matrice BAHDEN	12
2.4.2. La matrice SOJA7080	12
2.4.3. La matrice SOJA12	12
2.5. Exercice	14
3. Trois méthodes d'analyse en surfaces de tendances	15
3.1. La méthode MATHER	15
3.1.1. Le modèle polynomial	16
3.1.2. Principes d'interprétation	16
3.1.3. Utilisation du programme	16
3.1.4. Présentations des sorties	18
3.2. La méthode WRIGLEY	21
3.2.1. Le modèle LOGIT	21
3.2.2. Principes d'interprétation	22
3.2.3. Utilisation du programme et présentation des sorties	23
3.3. La méthode LEMAY	26
3.3.1. Réitérer une régression linéaire simple: méthode et interprétation	26
3.3.2. Utilisation du programme et présentation des sorties	27
4. CONCLUSION: Comment choisir l'une des trois méthodes ?	30
Eléments de bibliographie	31
ANNEXE: Le progiciel SURF	32
A.1. Configuration nécessaire	32
A.2. La disquette SURF	32
A.3. Installation et utilisation du progiciel	32
A.4. Saisir un fichier de données	32
A.5. Ajouter une ou plusieurs options de transformation de Z	34

Philippe WANIEZ, les données et le territoire: initiation au traitement informatique des données spatialisées.  
(ISBN 2-86912-013-8; 115 p.; 80 F).

Philippe WANIEZ, Violette CABOS, les données et le territoire: initiation à la numérisation pour la cartographie statistique.  
(ISBN 2-86912-015-4; 54p.; 40F).

Ces deux ouvrages sont en vente à l'adresse suivante:  
G.I.P. RECLUS, Maison de la Géographie, 17 rue Abbé de l'Epée  
34000 Montpellier Tel. 67 72 46 10

*D - Systèmes.*  
*Reprographie.*



REPROGRAPHIE INDUSTRIELLE  
EDITIONS - DUPLICATIONS -

MINIPARC N° 7-ZOLAD- 34100 MONTPELLIER  
07.61.02.60

ÉDITIONS DE L'ORSTOM

DIFFUSION :

70-74 route d'Aulnay - 93140 BONDY-FRANCE

ISBN 2-7099-0883-2

ISSN 0071-9021